T H E S I S

FOR THE DEGREE OF Ph.D.

"THE ESTIMATION OF STATISTICAL COEFFICIENTS FROM SAMPLES"

LEON SOLOMON, M.A.

CHAPTER ONE

INTRODUCTION

1.0 Statistics, it has been said, is the reduction of data. This definition fails, however, to bring out one of the most important and interesting aspects of modern statistical theory - the making of estimates. Let us consider a typical example. An investigator collects a body of data, consisting, say, of the heights, or other attribute, of some members of a community. He calculates a few quantities - such as a mean, a measure of dispersion, perhaps one of skewness - which enable him to apprehend the properties of the assemblage of figures. This is the reduction of data. But now curiosity or practical need poses a new question. What can be said regarding the heights of all the members of the community - not merely of those in respect of whom information is tabulated? Of course, nothing can be stated with certainty. The heights of the members who were measured may vary widely from those of the others. Nevertheless, intuition, or experience, suggests that at least a guess might be hazarded. One suspects, too, that in some instances a better guess is justified than in others. Can one, then, measure in some fashion the reliability of a guess? Questions such as these are the subject-matter of the Theory of Estimation.

This theory, as our example indicates, has nothing in common with the deductive problems frequently encountered in Pure Mathematics, such as the deduction of a particular geometrical theorem, given a few general axioms. On the contrary, Estimation is concerned with the realm of induction; with passing from the particular to the general; with the making of inferences regarding a population from the data of a sample thereof. By their nature, these inferences are uncertain. Perhaps it is the insufficient appreciation of this which has led to the well-known criticism that statistics "can be made to prove anything." It would be fairer to say that statistics <u>proves</u> nothing - but that it may suggest a great deal.

1.1 The Probability Distribution:- A probability distribution - the concept is one of the most useful in Mathematical Statistics - is defined thus :

Let a variate x be capable of assuming all values within some range, and let the probability that x should lie within the infinitesimal range $x_1 \pm \frac{1}{2}dx$, be $\varphi(x_1 | \theta_1, \cdots \theta_v)\,dx$, where the θ's are constants. Then $\varphi(x \mid \theta_1 \cdots \theta_v)$ is the probability distribution.

The coefficients $\theta_1, \theta_2, \cdots \theta_v$ once specified, we can evaluate the probability that an observation of x should lie within any given range; and, out of a large number n of random observations, we would expect the number within the range to be n times this probability.

R. A. Fisher ("On The Foundations of Theoretical Statistics," Philosophical Transactions, A, Vol. 222 (1922)) associates three distinct problems with a probability distribution, viz., those of

(a) specification

(b) estimation

(c) distribution.

The first concerns the choice of the mathematical form of the distribution φ most apposite to a given body of data. Though often of considerable difficulty, this problem is, as Fisher points out, essentially practical. The choice having been made, the actual data are regarded as a random sample of observations drawn from an infinite "universe" distributed in the form φ.

The problem of specification is assumed to be solved before that of estimation is begun. The latter is formally defined thus: given a random sample of n

observations $x_1, x_2, \cdots\cdots x_n$ of a population whose probability distribution is $\varphi(x \mid \theta_1 \cdot \theta_2 \cdot \cdots\cdots \theta_v)$ what functions of the observations $T_1(x_1, x_2, \cdots\cdots x_n)$, $\cdots\cdots\cdots T_v(x_1, x_2, \cdots x_n)$ give the best estimates of the unknown coefficients $\theta_1, \theta_2, \cdots\cdots \theta_v$ respectively?

The final task is to determine the sampling distribution of the selected functions $T_1, T_2, \cdots\cdots T_v$. While the analytical difficulties are often formidable, this is in principle a straightforward exercise in mathematical deduction.

1.2 Criteria For The Selection of Estimating Functions:-

In the preceding section, the phrase "the best estimate" was left unexplained. "Best," as applied to a question of induction, need not, of course, bear the same connotation as in the field of deductive logic, and different statisticians have adopted many different definitions. Nevertheless, the concensus of current opinion tends to regard consistency and minimum sampling variance as necessary attributes of a "good" estimate.

1.2.1 Consistency:- Let $T(x_1 \cdots\cdots x_n)$ be an estimate - or "statistic" - of the coefficient θ, in a sample of n, and let the probability that

$$|T(x_1 \cdots\cdots x_n) - \theta| < \epsilon$$

be $P_n(<\epsilon)$ where ϵ is any positive quantity, no matter how small. T is said to be consistent if

$$P_n(<\epsilon) \to 1 \quad \text{as} \quad n \to \infty.$$

In words, this definition provides little more than a warning that, if we wish to estimate θ, we must beware of estimating something other than θ. The condition of consistency is thus not very severe. It certainly does not determine T uniquely, and it leaves open the question which one of various consistent statistics is "best." Nevertheless, the postulate is of some value in itself. For instance, its rigorous application enables us to derive Sheppard's Corrections for grouped data (Fisher, loc. cit.).

1.2.2 Sampling Variance:- The n observations of a sample provide a statistic T which serves as an estimate of θ. If another n observations were made from the hypothetical infinite universe, we would doubtless obtain a different value as the estimate of θ. Imagine that an infinitely large number of samples, each of size n, were drawn, and the corresponding estimates of θ noted. This set of estimates would possess a mean, and a variance about the mean - which is termed the sampling variance of T. The stipulation is commonly made that the best statistic will be taken as that with the least sampling variance.

1.3 Methods of Estimation:-

1.3.1 The Method of Moments:- For a continuous probability distribution $\varphi(x \mid \theta_1 \cdots \theta_v)$ $[a \le x \le b]$

the r^{th} moment, about the origin, is

$$\mu_r = \int_a^b x^r \varphi(x \mid \theta_1 \cdots \theta_v)\, dx \qquad [r = 1, 2, \cdots]$$

The μ's are thus functions of $\theta_1, \theta_2, \cdots \theta_v$.

If the random observations are $x_1, x_2, \cdots x_n$, the r^{th} moment of the sample, about the origin, is

$$m_r = \sum_{i=1}^{n} x_i^r \qquad [r = 1, 2, \cdots]$$

Karl Pearson estimates the coefficients by equating population and sample moments -

$$\mu_r = m_r \qquad [r = 1, 2, \cdots v]$$

and solving the resulting equations. The estimates so obtained are obviously consistent.

<u>1.3.2</u> <u>Maximum Likelihood</u>:- A fundamental theorem concerning the existence of statistics of minimum sampling variance has been derived by R. A. Fisher. A very wide class of estimates is distributed normally (especially in infinitely large samples), and the variance V is then inversely proportionate to the size of the sample n. The theorem is that nV cannot be less than a certain quantity, which is independent of all methods of estimation. A statistic for which nV is equal to this quantity is called "efficient." Fisher demonstrated that the Method of Moments failed in general to yield efficient estimates, but that his new method succeeded.

This new method is developed in terms of "likelihood." For a distribution $\varphi(x \mid \theta_1, \theta_2, \cdots \theta_\nu)$ the probability that n observations should lie respectively within the small ranges $x_i \pm \frac{1}{2} dx_i \; [i = 1, 2, \cdots n]$ is

$$L \, dx_1 \, dx_2 \cdots dx_n = \prod_{i=1}^{n} \varphi(x_i \mid \theta_1 \cdots \theta_\nu) \, dx_i \; .$$

The quantity L is defined as the Likelihood. (A more descriptive name is simply "probability density") Fisher chooses as estimates of $\theta_1, \theta_2, \cdots \theta_\nu$ those values of the coefficients which make the likelihood a maximum - i.e., he solves the ν equations

$$\frac{\partial L}{\partial \theta_i} = 0 \qquad [i = 1, 2, \cdots \nu]$$

1.3.3 The postulate of maximum likelihood is - though the inventor might disagree - arbitrary. Its value is that the statistics so obtained have the important properties:

(i) they are consistent

(ii) they are efficient, in large samples

(iii) if "sufficient" statistics exist, they will be found by Fisher's method. Such a statistic contains the whole of the information which a finite sample can provide.

1.4 Estimation by an Unbiassed Statistic of Minimum Variance:- While the objectives of Maximum Likelihood are consistent and efficient estimates, they are attained indirectly - via the maximising of a certain function. An alternative approach is to assert initially that statistics should have these, or similar, attributes, and to omit any reference to likelihood. This was the starting point in the paper by A. C. Aitken and H. Silverstone ("On the Estimation of Statistical Parameters," Proceedings of The Royal Society of Edinburgh, A, Vol. 61) who adopted the criteria that an estimate should

(i) be unbiassed. That is, the expectation, over all samples of n, of an estimate of θ, should equal θ - even in finite samples.

(ii) have minimum sampling variance.

Why, one may ask, was unbiassedness selected, which is a more severe condition than mere consistency? The reason is simply one of convenience. Requirements of minimum variance and consistency do not lead to a solution. Those of minimum variance and unbiassedness do.

It is interesting to note some divergent opinions on the question of bias. R. A. Fisher considers it of little importance. Thus he says ("The Logic of Inductive Inference," Journal of The Royal Statistical Society, vol. 98 (1935)) "the consideration of bias need not detain us. With consistent estimates it must tend to zero.......

We can always adjust our estimate so as to make the bias absolutely zero, but this is not usually necessary....."
On the other hand, many writers do strive to avoid biassed statistics - witness the preference for $\frac{1}{n-1}\sum(x-\bar{x})^2$ to $\frac{1}{n}\sum(x-\bar{x})^2$ as an estimate of the variance of the normal curve of error.

<u>1.4.1</u> The study of the criteria of unbiassedness and minimum sampling variance is the main concern of this thesis. As we proceed, we shall discover both analogies with and differences from the Maximum Likelihood treatment, which shed a new and instructive light on the latter. We shall find that the theory we develop is independent of the size of the sample - whether large or small - and necessitates no assumptions whatever regarding the sampling distribution of statistics.

Interesting conclusions, too, emerge from the study of simultaneous estimation. One's first intuitive thought is that each statistic should be unbiassed and of minimum variance. Unfortunately, no such statistics exist for the estimation of two or more parameters. Alternative criteria, which are developed, are not free from surprising implications. Thus we shall set out to minimise a quantity -

generalised variance - but we shall succeed only in making it a weak minimum. The size of the sample must be indefinitely great, before we find statistics which make the minimum strong.

Such results as these suggest that the postulates of unbiassedness and of minimum variance are not unworthy of attention.

Chapter Two

LINEAR INTEGRAL EQUATIONS

2.0 The need to solve a linear integral equation often arises in statistical problems; for instance, in the problem of determining what probability distribution has a given moment generating function. In later chapters, a certain homogeneous linear integral equation will be encountered. We shall require to solve it, and to know whether the solution is unique. Since this particular equation is treated rather cursorily, and in some respects incompletely, in the text books, we shall establish in this chapter the various results which we invoke in the sequel.

Let us consider first the n-dimensional form of Fredholm's equation, viz.

$$u(x_1 \cdots x_n) = f(x_1 \cdots x_n) + \lambda \int_a^b \cdots \int_a^b K(x_1 \cdots x_n, \xi_1 \cdots \xi_n) u(\xi_1 \cdots \xi_n) \, d\xi_1 \cdots d\xi_n$$

which we abbreviate to

$$u(x') = f(x') + \lambda \int_a^b K(x' \cdot \xi') u(\xi') \, d\xi'$$

where x' denotes the vector $(x_1, x_2, \cdots x_n)$. Our problem is - given f and K, can a function (or functions) u be determined so that Fredholm's equation is satisfied for <u>every</u> value of $x' = (x_1 \cdots x_n)$ in the range $a \leq x_1 \leq b$, $a \leq x_2 \leq b$, ------- $a \leq x_n \leq b$?

(For compactness, we may express this set of n inequalities by $a \leq x' \leq b$).

The problem may be simplified by imposing the following restrictions (Some of them will be relaxed later)

(1) $f(x')$ is continuous with respect to each of its n arguments in the ranges $a \leq x' \leq b$.

(2) $K(x' \cdot \xi')$ is a finite continuous function of all its arguments throughout the 2n ranges $a \leq x' \leq b$, $a \leq \xi' \leq b$. Let us suppose that $|K(x' \cdot \xi'|$ is everywhere less than M, a finite quantity independent of x' and ξ'.

(3) $f(x')$ and $K(x' \cdot \xi')$ are real functions of real variables.

(4) λ is a real constant.

(5) a and b are finite.

2.1 The Associated Series: Corresponding to a given "kernel" $K(x' \cdot \xi')$ we introduce the following series.

$$D(\lambda) = 1 - \lambda \int_a^b K(\xi_1' \cdot \xi_1')\, d\xi_1' + \frac{\lambda^2}{2!}\int_a^b\int_a^b K\begin{pmatrix}\xi_1' & \xi_2' \\ \xi_1' & \xi_2'\end{pmatrix} d\xi_1' d\xi_2' - \cdots\cdots$$

$$D(x'|y';\lambda) = \lambda K(x' \cdot y') - \sum_{p=1}^{\infty} \frac{(-\lambda)^{p+1}}{p!} \int_a^b \cdots \int_a^b K\begin{pmatrix} x' \xi_1' \xi_2' \cdots \xi_p' \\ y' \xi_1' \xi_2' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'$$

$$D(x_1' \cdots x_r' | y_1' \cdots y_r' ; \lambda) = \lambda K\begin{pmatrix} x_1' \cdots\cdots x_r' \\ y_1' \cdots\cdots y_r' \end{pmatrix}$$

$$- \sum_{p=1}^{\infty} \frac{(-\lambda)^{p+1}}{p!} \int_a^b \cdots \int_a^b K\begin{pmatrix} x_1' \cdots x_r' \xi_1' \cdots \xi_p' \\ y_1' \cdots y_r' \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'$$

$[r = 2, 3, 4, \cdots]$

where
$$K\begin{pmatrix} x_1' \cdots x_r' \\ y_1' \cdots y_r' \end{pmatrix} = \begin{vmatrix} K(x_1' \cdot y_1') & K(x_1' \cdot y_2') & \cdots\cdots & K(x_1' \cdot y_r') \\ K(x_2' \cdot y_1') & K(x_2' \cdot y_2') & \cdots\cdots & K(x_2' \cdot y_r') \\ \vdots & \vdots & & \vdots \\ K(x_r' \cdot y_1') & K(x_r' \cdot y_2') & \cdots\cdots & K(x_r' \cdot y_r') \end{vmatrix}$$

The second series, D $(x'|y';\lambda)$ is of course a particular case of the third series, viz., the case when r = 1.

2.1.1. Convergence of The Associated Series: Since $K(x'\cdot y')$ is bounded, we may apply Hadamard's lemma on the maximum of a determinant (Whittaker and Watson's Modern Analysis, 4th edition, p. 212), whence

$$\left| K\begin{pmatrix} x_1' \cdots & x_r' \\ y_1' \cdots & y_r' \end{pmatrix} \right| < M^r r^{r/2}$$

Now consider D (λ). If we write this as $1 + \sum_1^\infty a_p \lambda^p / p!$, then

$$|a_p| = \left| \int_a^b \cdots \int_a^b K\begin{pmatrix} \xi_1' \cdots & \xi_p' \\ \xi_1' \cdots & \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p' \right| < M^p p^{p/2} \int_a^b \cdots \int_a^b d\xi_1' \cdots d\xi_p'$$

$$\therefore \; |a_p| < M^p p^{p/2} (b-a)^{np}.$$

Write $c_p = M^p p^{p/2} (b-a)^{np} / p!$

$$\therefore \; \frac{c_{p+1}}{c_p} = \frac{M(b-a)^n}{p+1} \; \frac{(p+1)^{\frac{p+1}{2}}}{p^{p/2}}$$

$$= M(b-a)^n \left\{ \left(1 + \frac{1}{p}\right)^p \right\}^{1/2} \frac{1}{(p+1)^{1/2}} \to 0 \text{ as } p \to \infty$$

since $(1+1/p)^p \to e$.

Hence the series $\sum_{p=1}^{\infty} c_p \lambda^p$ is absolutely convergent for all values of λ; the series for D(λ) is therefore absolutely convergent for all values of λ and representx an integral function of λ.

With regard to $D(x_1' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda)$ we note first that by Hadamard's lemma

$$\left| K\begin{pmatrix} x_1' \cdots x_r' & \xi_1' \cdots \xi_p' \\ y_1' \cdots y_r' & \xi_1' \cdots \xi_p' \end{pmatrix} \right| < M^{r+p} (r+p)^{\frac{r+p}{2}}$$

whence

$$\left| \int_a^b \cdots \int_a^b K\begin{pmatrix} x_1' \cdots x_r' & \xi_1' \cdots \xi_p' \\ y_1' \cdots y_r' & \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p' \right| < M^{r+p} (r+p)^{\frac{r+p}{2}} (b-a)^{np}$$
$$= c_p \quad \text{say}.$$

Applying Weierstrass' test, we consider the series $\sum_0^\infty \frac{|\lambda|^{p+1} c_p}{p!}$ each term of which is greater than the absolute value of the corresponding term in $D(x_1' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda)$. The ratio of the $(p+1)$ st. term to the p^{th} term of this dominant series is

$$|\lambda| M (b-a)^n \left\{ \left(1 + \frac{1}{r+p}\right)^{r+p} \right\}^{\frac{1}{2}} \left(1 + \frac{r}{p+1}\right)^{\frac{1}{2}} \frac{1}{(p+1)^{1/2}}$$

$$\longrightarrow 0 \quad \text{as } p \to \infty.$$

It follows that $D(x_1' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda)$ is absolutely convergent for all finite values of λ, and is uniformly convergent for all values of the vectors $x_1' \cdots x_r'$, $y_1' \cdots y_r'$ within the ranges
$a \le x_i' \le b$, $a \le y_i' \le b$ $[i = 1, 2, \cdots r;\ r = 1, 2, \cdots]$
Further, since $K(x' \cdot y')$ is a continuous function of all its variables, it is readily seen that each term in this series is a continuoue function of all its variables.

The same results obviously hold for $D(x'/y'; \lambda)$ since this series is obtained by putting $r = 1$ in the foregoing.

2.1.2. The Principle of Reciprocity: Certain important identities connect the associated series with one another. The simplest of them, termed by Volterra "the principle of reciprocity," is derived as follows. Since $D(x'/y'; \lambda)$ is absolutely convergent for all finite values of λ, rearrange the terms by picking out the coefficients of $K(x' \cdot y')$ and writing these first. There results the equation

$$D(x'/y'; \lambda) = \lambda D(\lambda) K(x' \cdot y') - \sum_{p=1}^{\infty} \frac{(-\lambda)^{p+1}}{p!} Q_p(x' \cdot y') \qquad (1)$$

where

$$Q_p(x' \cdot y') = \int_a^b \cdots \int_a^b \begin{vmatrix} 0 & K(x' \cdot \xi_1') & K(x' \cdot \xi_2') \cdots\cdots & K(x' \cdot \xi_p') \\ K(\xi_1' \cdot y') & K(\xi_1' \cdot \xi_1') & & K(\xi_1' \cdot \xi_p') \\ \vdots & \vdots & & \vdots \\ K(\xi_p' \cdot y') & K(\xi_p' \cdot \xi_1') & & K(\xi_p' \cdot \xi_p') \end{vmatrix} d\xi_1' \cdots d\xi_p'$$

Expanding in minors of the first column, we obtain Qp as the integral of a sum of p determinants. The first term in the sum is

$$-\int_a^b\cdots\int_a^b \begin{vmatrix} K(x'\cdot\xi_1') & K(x'\cdot\xi_2') & \cdots\cdots & K(x'\cdot\xi_p') \\ K(\xi_2'\cdot\xi_1') & K(\xi_2'\cdot\xi_2') & \cdots\cdots & K(\xi_2'\cdot\xi_p') \\ \vdots & \vdots & & \vdots \\ K(\xi_p'\cdot\xi_1') & K(\xi_p'\cdot\xi_2') & \cdots\cdots & K(\xi_p'\cdot\xi_p') \end{vmatrix} K(\xi_1'\cdot y')\,d\xi_1'\cdots d\xi_p'$$

$$= -\int_a^b\cdots\int_a^b \begin{vmatrix} K(x'\cdot\xi') & K(x'\cdot\xi_1') & & K(x'\cdot\xi_{p-1}') \\ K(\xi_1'\cdot\xi') & K(\xi_1'\cdot\xi_1') & & K(\xi_1'\cdot\xi_{p-1}') \\ & & & \\ K(\xi_{p-1}'\cdot\xi') & K(\xi_{p-1}'\cdot\xi_1') & & K(\xi_{p-1}'\cdot\xi_{p-1}') \end{vmatrix} K(\xi'\cdot y')\,d\xi'\,d\xi_1'\cdots d\xi_{p-1}'$$

on changing ξ_1' to ξ', ξ_2' to ξ_1', $\cdots\cdots$ ξ_p' to ξ_{p-1}', in the variables of integration.

The second term is

$$\int_a^b\cdots\int_a^b \begin{vmatrix} K(x'\cdot\xi_1') & K(x'\cdot\xi_2') & & K(x'\cdot\xi_p') \\ K(\xi_1'\cdot\xi_1') & K(\xi_1'\cdot\xi_2') & & K(\xi_1'\cdot\xi_p') \\ K(\xi_3'\cdot\xi_1') & K(\xi_3'\cdot\xi_2') & & K(\xi_3'\cdot\xi_p') \\ & & & \\ K(\xi_p'\cdot\xi_1') & K(\xi_p'\cdot\xi_2') & & K(\xi_p'\cdot\xi_p') \end{vmatrix} K(\xi_2'\cdot y')\,d\xi_1'\,d\xi_2'\cdots d\xi_p'.$$

If we write, for the variables of integration, ξ' in place of ξ_2'; ξ_2' in place of ξ_3'; $\cdots\cdots$ ξ_{p-1}' in place of ξ_p'; and if we moreover interchange the first two columns of the determinant, this second term becomes

$$-\int_a^b\cdots\int_a^b \begin{vmatrix} K(x'\cdot\xi') & K(x'\cdot\xi_1') & \cdots & K(x'\cdot\xi_{p-1}') \\ K(\xi_1'\cdot\xi') & K(\xi_1'\cdot\xi_1') & \cdots & K(\xi_1'\cdot\xi_{p-1}') \\ \vdots & \vdots & & \vdots \\ K(\xi_{p-1}'\cdot\xi') & K(\xi_{p-1}'\cdot\xi_1') & \cdots & K(\xi_{p-1}'\cdot\xi_{p-1}') \end{vmatrix} K(\xi'\cdot y')\,d\xi'd\xi_1'\cdots d\xi_{p-1}'$$

The second term in the expansion of $Qp(x'\cdot y')$ is therefore equal to the first term. Similarly, all the other $(p-2)$ terms are equal to the first term, so that equation (1) may be written

$$D(x'|y';\lambda) = \lambda D(\lambda)K(x'\cdot y') + \sum_{p=1}^{\infty}\frac{(-\lambda)^{p+1}}{(p-1)!}\int_a^b K(\xi'\cdot y')d\xi'\int_a^b\cdots\int_a^b K\begin{pmatrix} x'\xi_1'\cdots\xi_{p-1}' \\ y'\xi_1'\cdots\xi_{p-1}'\end{pmatrix}d\xi_1'\cdots d\xi_{p-1}'$$

Recalling the original value of $D(x'|y';\lambda)$ this becomes $D(x'|y';\lambda) = \lambda D(\lambda)K(x'\cdot y') + \lambda\int_a^b K(\xi'\cdot y')D(x'|\xi';\lambda)\,d\xi'$ (2).

Similarly, by expanding Qp in minors of the first row, we obtain

$$D(x'|y';\lambda) = \lambda D(\lambda)K(x'\cdot y') + \lambda\int_a^b K(x'\cdot\xi')D(\xi'|y';\lambda)\,d\xi' \qquad (3)$$

Formulae (2) and (3) constitute the Principle of Reciprocity.

Corresponding formulae involving the $D(x_1'\cdots x_r'|y_1'\cdots y_r';\lambda)$ exist. Thus, expanding in minors of the first row, we have

$$\frac{1}{p!}\int_a^b\cdots\int_a^b K\begin{pmatrix} x_1' \cdots\cdots x_r' & \xi_1' \cdots & \xi_p' \\ y_1' \cdots\cdots y_r' & \xi_1' \cdots & \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'$$

$$= \frac{1}{p!} K(x_1' \cdot y_1') \int_a^b\cdots\int_a^b K\begin{pmatrix} x_2' \cdots & x_r' & \xi_1' \cdots \xi_p' \\ y_2' \cdots & y_r' & \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'$$

$$- \frac{1}{p!} K(x_1' \cdot y_2') \int_a^b\cdots\int_a^b K\begin{pmatrix} x_2' \, x_3' \cdots x_r' & \xi_1' \cdots \xi_p' \\ y_1' \, y_3' \cdots y_r' & \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'$$

$+ (r-2)$ similar terms

$$+ \frac{(-1)^r}{p!}\int_a^b\cdots\int_a^b K(x_1' \cdot \xi_1') K\begin{pmatrix} x_2' \cdots & x_r' & \xi_1' \cdots & \xi_p' \\ y_1' \cdots & y_{r-1}' & y_r' \cdots & \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'$$

$-(p - 1)$ similar terms.

Now the last p terms are equal to one another, with alternate signs (since they contain the same determinant, with the columns in different order). They may therefore be replaced by the single term

$$\frac{(-1)^r}{(p-1)!}\int_a^b\cdots\int_a^b K(x_1' \cdot \tau') K\begin{pmatrix} x_2' \cdots\cdots x_r' & \tau' \, \xi_2' \cdots \xi_p' \\ y_1' \cdots\cdots y_{r-1}' & y_r' \, \xi_2' \cdots \xi_p' \end{pmatrix} d\tau' d\xi_2' \cdots d\xi_p'$$

$$= \frac{-1}{(p-1)!}\int_a^b\cdots\int_a^b K(x_1' \cdot \tau') K\begin{pmatrix} \tau \;\; x_2' \cdots & x_r' & \xi_1' \cdots \xi_{p-1}' \\ y_1' \cdots\cdots & y_r' & \xi_1' \cdots \xi_{p-1}' \end{pmatrix} d\tau' d\xi_1' \cdots d\xi_{p-1}'$$

Using this expression in the original series for $D(x' \cdots x_r' \,|\, y_1' \cdots y_r' \,;\, \lambda)$, viz.

$$\lambda K\begin{pmatrix} x_1' \cdots & x_r' \\ y_1' \cdots & y_r' \end{pmatrix} - \sum_{p=1}^{\infty} \frac{(-\lambda)^{p+1}}{p!}\int_a^b\cdots\int_a^b K\begin{pmatrix} x_1' \cdots x_r' & \xi_1' \cdots \xi_p' \\ y_1' \cdots y_r' & \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'$$

we obtain

$$
\begin{aligned}
&D(x_1' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda) \\
&= K(x_1' \cdot y_1')\left[\lambda K\begin{pmatrix} x_2' \cdots x_r' \\ y_2' \cdots y_r' \end{pmatrix} - \sum_{p=1}^{\infty} \frac{(-\lambda)^{p+1}}{p!} \int_a^b \cdots \int_a^b K\begin{pmatrix} x_2' \cdots x_r' \xi_1' \cdots \xi_p' \\ y_2' \cdots y_r' \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'\right] \\
&- \cdots\cdots \\
&+ (-1)^{r-1} K(x_1' y_r')\left[\lambda K\begin{pmatrix} x_2' \cdots x_r' \\ y_1' \cdots y_{r-1}' \end{pmatrix} - \sum_{p=1}^{\infty} \frac{(-\lambda)^{p+1}}{p!} \int_a^b \cdots \int_a^b K\begin{pmatrix} x_2' \cdots x_r' & \xi_1' \cdots \xi_p' \\ y_1' \cdots y_{r-1}' & \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'\right] \\
&+ \int_a^b K(x_1' \cdot \tau') d\tau' \sum_{p=1}^{\infty} \frac{(-\lambda)^{p+1}}{(p-1)!} \int_a^b \cdots \int_a^b K\begin{pmatrix} \tau' x_2' \cdots x_r' & \xi_1' \cdots \xi_{p-1}' \\ y_1' \cdots y_r' & \xi_1' \cdots \xi_{p-1}' \end{pmatrix} d\xi_1' \cdots d\xi_{p-1}'
\end{aligned}
$$

(The rearrangement of terms is legitimate since the series is absolutely and uniformly convergent)

Recalling the original expression for the $D(x_1' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda)$ this last identity may be rewritten in the form

$$
\begin{aligned}
D(x_1' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda) &= K(x_1' \cdot y_1') D(x_2' \cdots x_r' \mid y_2' \cdots y_r' ; \lambda) \\
&- K(x_1' \cdot y_2') D(x_2' \cdots x_r' \mid y_1' y_3' \cdots y_r' ; \lambda) + \cdots\cdots \\
&+ (-1)^{r-1} K(x_1' \cdot y_r') D(x_2' \cdots x_r' \mid y_1' \cdots y_{r-1}' ; \lambda) \\
&+ \lambda \int_a^b K(x_1' \cdot \tau') D(\tau' x_2' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda) d\tau' \qquad (4)
\end{aligned}
$$

If we had expanded our determinants in minors of the first column instead of the first row, we would have deduced

$$
\begin{aligned}
D(x_1' \cdots x_r' \mid y_1' \cdots y_r' ; \lambda) &= K(x_1' \cdot y_1') D(x_2' \cdots x_r' \mid y_2' \cdots y_r' ; \lambda) \\
&- K(x_2' \cdot y_1') D(x_1' x_3' \cdots x_r' \mid y_2' \cdots y_r' ; \lambda) + \cdots\cdots \\
&+ (-1)^{r-1} K(x_r' \cdot y_1') D(x_1' \cdots x_{r-1}' \mid y_2' \cdots y_r' ; \lambda) \\
&+ \lambda \int_a^b K(\tau' \cdot y_1') D(x_1' \cdots x_r' \mid \tau' y_2' \cdots y_r' ; \lambda) d\tau' \qquad (5)
\end{aligned}
$$

The formulae (4) and (5) have been termed the Generalised Principle of Reciprocity.

2.1.3. The Iterated Functions of The Kernel $K(x' \cdot y')$

In addition to the series studied above, the "iterated functions" $K_1(x' \cdot y')$, $K_2(x' \cdot y')$, ---- are of importance. They are defined by

$$K_p(x' \cdot y') = \int_a^b K_{p-1}(x' \xi') K_1(\xi' \cdot y')\, d\xi' \qquad (p = 2, 3, \cdots)$$
$$K_1(x' \cdot y') = K(x' \cdot y')$$

We can readily show that the reciprocity

$$\begin{aligned} K_p(x' \cdot y') &= \int_a^b K_{p-1}(x' \cdot \xi') K_1(\xi' \cdot y')\, d\xi' \\ &= \int_a^b K_1(x' \cdot \xi') K_{p-1}(\xi' y')\, d\xi' \end{aligned}$$

subsists $(p = 1, 2, 3 \cdots)$. For suppose it is true for the value $p - 1$. Then

$$\begin{aligned} K_p(x' \cdot y') &= \int_a^b K_{p-1}(x' \cdot \xi') K_1(\xi' \cdot y')\, d\xi' \\ &= \int_a^b K_1(\xi' \cdot y')\, d\xi' \int_a^b K_{p-2}(x' \cdot \eta') K_1(\eta' \cdot \xi')\, d\eta' \\ &= \int_a^b K_1(\xi' \cdot y')\, d\xi' \int_a^b K_1(x' \cdot \eta') K_{p-2}(\eta' \cdot \xi')\, d\eta' \end{aligned}$$

by hypothesis

$$\begin{aligned} \therefore K_p(x' \cdot y') &= \int_a^b K_1(x' \cdot \eta')\, d\eta' \int_a^b K_{p-2}(\eta' \cdot \xi') K_1(\xi' \cdot y')\, d\xi' \\ &= \int_a^b K_1(x' \cdot \eta') K_{p-1}(\eta' \cdot y')\, d\eta' . \end{aligned}$$

Since the reciprocity is obviously true for p = 2, it is true in general, by induction.

If K $(x' \cdot y')$ is bounded, all the iterated functions are bounded. For suppose this is so for the (p - 1)st. iterated function: suppose in fact that $|K_{p-1}(x' \cdot y')| < M_{(p-1)}$ where $M_{(p-1)}$ is finite and independent of the 2 n variables $x' \cdot y'$. Then

$$|K_p(x' \cdot y')| = \left| \int_a^b K_{p-1}(x' \cdot \xi') K_1(\xi' \cdot y')\, d\xi' \right| < M \cdot M_{(p-1)} (b-a)^n$$

where M is the upper bound of K ($= K_1$). The truth of our assertion therefore follows by induction. It follows from the last inequality, too, that we can take

$$M_{(p-1)} = M^{p-1} (b-a)^{(p-2)n} \qquad (p = 2, 3, \cdots)$$

whence

$$|K_p(x' \cdot y')| < M^p (b-a)^{(p-1)n} .$$

2.2 The Solution of Fredholm's Equation: We are now in a position to tackle our initial problem, viz., the solution of the integral equation

$$u(x') = f(x') + \lambda \int_a^b K(x' \cdot \xi') u(\xi')\, d\xi' .$$

Let us multiply by $D(y'|x';\lambda)$ and integrate with respect to the n variables x' between a and b. We have

$$\int_a^b u(x') D(y'|x';\lambda)\, dx' = \int_a^b f(x') D(y'|x';\lambda)\, dx' + \lambda \int_a^b D(y'|x';\lambda) dx' \int_a^b K(x' \cdot \xi') u(\xi')\, d\xi'$$

$$= \int_a^b f(x') D(y'|x';\lambda)\, dx' + \lambda \int_a^b u(\xi')\, d\xi' \int_a^b K(x' \cdot \xi') D(y'|x';\lambda)\, dx'$$

(the inversion of the integrals being legitimate for our continuous, bounded functions)

$$= \int_a^b f(x') D(y'|x';\lambda)\, dx' + \int_a^b u(\xi') d\xi' \{ D(y'|\xi';\lambda) - \lambda D(\lambda) K(y' \cdot \xi') \}$$

by the Principle of Reciprocity. Hence

$$\lambda D(\lambda) \int_a^b u(\xi') K(y' \cdot \xi')\, d\xi' = \int_a^b f(x')\, D(y'|x';\lambda)\, dx'$$

or, from the original equation,

$$D(\lambda)\{u(y') - f(y')\} = \int_a^b f(x')\, D(y'|x';\lambda)\, dx'$$

Therefore, if $D(\lambda) \neq 0$

$$u(y') = f(y') + \frac{1}{D(\lambda)}\int_a^b f(x')D(y'|x';\lambda)dx' \qquad (6)$$

If, then, Fredholm's equation has a solution, it can be none other than (6). Substituting in the original equation, we see that this is in fact a solution, which is, therefore, unique and continuous (by 2.1.1), provided $D(\lambda) \neq 0$.

2.2.1 Solution by Successive Approximations: Let us take as a first approximation to the solution of Fredholm's equation

$$u_1(x') = f(x').$$

Substituting in the equation, the second approximation is

$$u_2(x') = f(x') + \lambda\int_a^b K(x'\cdot\xi')f(\xi')d\xi'$$

The third approximation is consequently

$$u_3(x') = f(x') + \lambda\int_a^b K(x'\cdot\xi')f(\xi')d\xi' + \lambda^2\int_a^b\int_c^b K(x'\cdot\xi_1')K(\xi_1'\cdot\xi')f(\xi')d\xi'd\xi_1'$$

or, in terms of the iterated functions of K,

$$u_3(x') = f(x') + \lambda\int_a^b K_1(x'\cdot\xi')f(\xi')d\xi' + \lambda^2\int_a^b K_2(x'\cdot\xi')f(\xi')d\xi'$$

Continuing the process, we find that the pth approximation is

$$u_p(x') = f(x') + \sum_{t=1}^{p-1}\lambda^t\int_a^b K_t(x'\cdot\xi')f(\xi')d\xi'$$

Let the upper bound of $|f(\xi')|$ be M'. Then, using the inequality established in 2.1.3, the modulus of the general term of this series does not exceed

$$|\lambda|^t M^t M' (b-a)^{(t-1)n}$$

Hence $u_p(x')$ converges uniformly (as $p \to \infty$) when

$$|\lambda| < M^{-1}(b-a)^{-n}$$

to $u_{(\infty)}(x')$ say; and, by actual substitution, $u_{(\infty)}(x')$ satisfies the integral equation.

Since, as we know, Fredholm's equation has only one solution if $D(\lambda) \neq 0$, it follows from 2.2 that

$$\frac{D(y'|x';\lambda)}{D(\lambda)} = \sum_{p=1}^{\infty} \lambda^p K_p(y' \cdot x')$$

provided $|\lambda| < M^{-1}(b-a)^{-n}$.

2.2.2 Taking $f(x') \equiv 0$, we deduce from 2.2 that when $D(\lambda) \neq 0$, the only continuous solution of the homogeneous equation

$$u(x') = \lambda \int_a^b K(x' \cdot \xi') u(\xi') d\xi'$$

is $u \equiv 0$.

2.3 Fredholm's Homogeneoue Integral Equation:

A theorem of fundamental importance is that the equation

$$u(x') = \lambda \int_a^b K(x' \cdot \xi') u(\xi') d\xi'$$

admits non-zero solutions, when λ is a root of $D(\lambda) = 0$.

By definition

$$D(\lambda) = 1 - \sum_{p=1}^{\infty} \frac{(-\lambda)^p}{p!} \int_a^b \cdots \int_a^b K\begin{pmatrix} \xi_1' \cdots \xi_p' \\ \xi_1' \cdots \xi_p' \end{pmatrix} d\xi_1' \cdots d\xi_p'.$$

Differentiating with respect to λ (a legitimate procedure, by 2.1.1), we have

$$D'(\lambda) = -\frac{1}{\lambda}\int_a^b D(\xi_1'|\xi_1';\lambda)\,d\xi_1'$$

$$D''(\lambda) = \frac{1}{\lambda}\int_a^b\int_a^b D(\xi_1'\xi_2'|\xi_1'\xi_2';\lambda)\,d\xi_1'\,d\xi_2'$$

$$D^{(p)}(\lambda) = \frac{(-1)^p}{\lambda}\int_a^b\cdots\int_a^b D(\xi_1'\cdots\xi_p'|\xi_1'\cdots\xi_p';\lambda)\,d\xi_1'\cdots d\xi_p'$$

The equation $D(\lambda) = 0$ has roots of finite multiplicity. Let λ be a root of multiplicity m, so that $D(\lambda) = D'(\lambda) = \cdots = D^{(m-1)}(\lambda) = 0$, but $D^{(m)}(\lambda) \neq 0$. So $D(\xi_1'\cdots\xi_m'|\xi_1'\cdots\xi_m';\lambda) \not\equiv 0$. Whatever the value of λ, there is therefore always one at least of the quantities $D(\xi_1'\cdots\xi_m'|\xi_1'\cdots\xi_m';\lambda)$ not identically zero. For our particular root λ, let p be the smallest integer for which $D(\xi_1'\cdots\xi_p'|\eta_1'\cdots\eta_p';\lambda)$ is not identically zero. (Therefote $p \leq m$, the multiplicity). The quantities $D(\xi_1'\cdots\xi_{p-1}'|\eta_1'\cdots\eta_{p-1}';\lambda)$ being identically zero, the generalised principle of reciprocity reduces to

$$D(\xi_1'\cdots\xi_p'|\eta_1'\cdots\eta_p';\lambda) = \lambda\int_a^b K(\xi_1'\tau')\,D(\tau'\xi_2'\cdots\xi_p'|\eta_1'\cdots\eta_p';\lambda)\,d\tau' \tag{4'}$$

Now put $D(\tau'\xi_2'\cdots\xi_p'|\eta_1'\cdots\eta_p';\lambda) = u_1(\tau')$

so that $\qquad D(\xi_1'\xi_2'\cdots\xi_p'|\eta_1'\cdots\eta_p';\lambda) = u_1(\xi_1')$

and we have the result that $u_1(\tau')$ is a solution of the homogeneous equation:

$$u_1(\xi_1') = \lambda\int_a^b K(\xi_1'\tau')\,u_1(\tau')\,d\tau'$$

Similarly

$$u_2(\tau') = D(\xi_1'\ \tau'\ \xi_3' \cdots \xi_p' \mid \eta_1' \cdots \eta_p'\ ;\lambda)$$
$$u_3(\tau') = D(\xi_1'\ \xi_2'\ \tau' \cdots \xi_p' \mid \eta_1' \cdots \eta_p'\ ;\lambda)$$

$$u_p(\tau') = D(\xi_1' \cdots \xi_{p-1}'\ \tau' \mid \eta_1' \cdots \eta_p'\ ;\lambda)$$

are solutions of this same equation. Also the linear combination

$$A_1 u_1 + A_2 u_2 + \cdots + A_p u_p$$

is a solution, where the A's are constants. The ξ's and η's which enter the expressions for the u's may be given any constant values)

The functions $u_1, \cdots u_p$ are described as the "fundamental solutions" corresponding to a given root λ of $D(\lambda) = 0$. The number of them, p, is termed the <u>index</u> of the root λ. From the foregoing, we see that the index is not greater than the multiplicity of the root. (Another proof of this important result is given later, in 2.5.4)

We shall now prove that the fundamental solutions are <u>linearly independent</u>. For, if we put $\xi_1' = \xi_2'$ (i.e., if we put each of the n variables in ξ_1' equal to the corresponding variable in ξ_2') in equation (4), we find that

$$\int_a^b K(\xi_1'\cdot\tau')\,D(\tau'\cdot\xi_1'\ \xi_3' \cdots \xi_p' \mid \eta_1' \cdots \eta_p'\ ;\lambda)\,d\tau'$$
$$= \frac{1}{\lambda} D(\xi_1'\cdot\xi_1'\ \xi_3' \cdots \xi_p' \mid \eta_1' \cdots \eta_p'\ ;\lambda) = 0$$

since the right hand side is the sum of determinants with two identical rows. That is, (since

$$D(\tau' \cdot \xi_1' \cdot \xi_3' \cdots \xi_p' \mid \eta_1' \cdots \eta_p' ; \lambda) = -D(\xi_1' \tau' \xi_3' \cdots \xi_p' \mid \eta_1' \cdots \eta_p' ; \lambda) = -u_2(\tau')$$

$$\int_a^b K(\xi_1' \cdot \tau') u_2(\tau') \, d\tau' = 0 ;$$

similarly

$\int_a^b K(\xi_i' \cdot \tau') u_j(\tau') \, d\tau' = 0$, provided $i \neq j$,

and, of course, $\int_a^b K(\xi_i' \cdot \tau') u_i(\tau') \, d\tau' = u_i(\xi_i')$.

Let A_1 --- A_p be constants (complex if necessary) and let A_i' denote the complex conjugate of A_i.

Consider

$$I = \lambda \int_a^b \left\{ \sum_{i=1}^p A_i u_i(\tau') \right\} \left\{ \sum_{i=1}^p A_i' K(\xi_i' \cdot \tau') \right\} d\tau'.$$

By the orthogonal property just established

$$I = \lambda \sum_{i=1}^p A_i A_i' \int_a^b K(\xi_i' \cdot \tau') u_i(\tau') \, d\tau'$$

$$= \lambda \sum_{i=1}^p A_i A_i' u_i(\xi_i')$$

But $u_1(\xi_1') = u_2(\xi_2') = \cdots = u_p(\xi_p') = D(\xi_1' \cdot \xi_2' \cdots \xi_p' / \eta_1' \cdots \eta_p' ; \lambda)$

Therefore

$$I = D(\xi_1' \cdots \xi_p' / \eta_1' \cdots \eta_p' ; \lambda) \sum_{i=1}^p A_i A_i' .$$

Now if the fundamental solutions are connected by any linear relation $\sum A_i u_i = 0$, then I is by definition zero. Therefore $\sum_{i=1}^p A_i A_i' = 0$.

Since this is a sum of squares, it requires

$A_i = 0 \quad (i = 1, 2, \cdots p)$. The fundamental solutions are therefore linearly independent.

2.4 Symmetric Kernels: Suppose that $K(x' \cdot y') = K(y' \cdot x')$. Then all the iterated functions are also symmetrical, and none is identically zero.

To prove the first assertion, assume that it is true for the p th iterated function. Then

$$K_{p+1}(x' \cdot y') = \int_a^b K_p(x' \cdot \xi') K(\xi' \cdot y') \, d\xi'$$

$$= \int_a^b K_p(\xi' \cdot x') K(y' \cdot \xi') \, d\xi'$$

(since K, Kp are symmetrical, by hypothesis)

$$= \int_a^b K(\xi' \cdot x') K_p(y' \cdot \xi') \, d\xi'$$

(on writing the expression in full as a p n fold integral, and reversing the order of integration)

$$= K_{p+1}(y' \cdot x')$$

Since K_1 is symmetrical by definition, the result follows by induction.

Suppose, if possible, that the second assertion is false, and that Kp, say, is the first iterated function which vanishes identically. Then K_{p+1} is also identically zero. Let p or $p+1$, whichever is an even integer, equal $2q$. Therefore

$$0 \equiv K_{2q}(x' \cdot y') \equiv \int_a^b K_{2q-1}(x' \cdot \xi') K(\xi' \cdot y') \, d\xi'$$

$$\equiv \int_a^b K_q(x' \cdot \xi') K_q(\xi' \cdot y') \, d\xi'$$

(on writing the expression in full as a $(2q-1) \times$ n fold integral, and changing the order of integration)

$$\equiv \int_a^b K_q(x'.\xi')\,K_q(y'.\xi')\,d\xi'$$

since all the iterated functions are symmetrical. In particular, when the vectors x', y' are equal

$$0 \equiv K_{2q}(x'.x') = \int_a^b \{K_q(x'.\xi')\}^2\,d\xi'$$

So $K_q \equiv 0$, which contradicts our hypothesis that K p is the first identically zero iterated function. Hence none of the iterated functions of a symmetrical kernel is zero.

When a kernel $K(x'.y')$ is symmetric, the roots of $D(\lambda) = 0$ are called the characteristic numbers of the kernel, and the corresponding non-zero continuous solutions of the homogeneoue integral equation are called characteristic functions.

2.5 Characteristic Numbers and Functions:

2.5.1 Schmidt's Theorem: A symmetric kernel has at least one characteristic number.

Let the kernel be $K(x'.y') = K(y'.x')$. From 2.2.1 we have

$$\frac{D(y'|x';\lambda)}{D(\lambda)} = \lambda K_1(y'.x') + \lambda^2 K_2(y'.x') + \cdots$$

Now $D(y'|x';\lambda)$ and $D(\lambda)$ are convergent for all finite values of λ, and the former is uniformly convergent for all values of the 2 n variables in the ranges $a \le x' \le b;\ a \le y' \le b$. Hence if, for some particular value λ, $D(y'|x';\lambda)/D(\lambda)$ is not uniformly convergent, that value must be a root of $D(\lambda) = 0$.

Assume that $D(y'|x';\lambda)/D(\lambda)$ is uniformly convergent for all values of λ. Then

$$\lambda^2 K_2(x'\cdot x') + \lambda^3 K_3(x'\cdot x') + \cdots$$

is uniformly convergent for all values of the vector x' and for all finite values of λ. As its terms are continuous, it can be integrated term by term with respect to the n variables x'. Let $\mu_p = \int_a^b K_p(x'\cdot x')\,dx'$. Then

$$\lambda^2 \mu_2 + \lambda^3 \mu_3 + \cdots\cdots$$

converges for all finite values of λ. As it is a power series, it is necessarily absolutely convergent. Therefore a part of the series, say

$$\lambda^2 \mu_2 + \lambda^4 \mu_4 + \lambda^6 \mu_6 + \cdots\cdots$$

converges for all finite values of λ.

Since K is symmetric

$$\begin{aligned}\mu_{p+q} &= \int_a^b K_{p+q}(x'\cdot x')\,dx' \\ &= \int_a^b\int_a^b K_p(x'\cdot \xi')\,K_q(\xi'\cdot x')\,dx'\,d\xi' \\ &= \int_a^b\int_a^b K_p(x'\cdot \xi')\,K_q(x'\cdot \xi')\,dx'\,d\xi'\end{aligned}$$

and $$\mu_{2p} = \int_a^b\int_a^b \{K_p(x'\cdot\xi')\}^2\,dx'\,d\xi' \qquad (7)$$

Let α, β be arbitrary real parameters, whence

$$\int_a^b\int_a^b \{\alpha K_{p+1}(x'\cdot\xi') + \beta K_{p-1}(x'\cdot\xi')\}^2\,dx'\,d\xi' \geq 0.$$

Therefore $\alpha^2 \mu_{2p+2} + 2\alpha\beta\,\mu_{2p} + \beta^2 \mu_{2p-2} \geq 0$

and $\mu_{2p+2}\cdot\mu_{2p-2} \geq \mu_{2p}^2.$

Since K is symmetric, none of the iterated functions is identically zero. So the μ_{2p}'s are (from (7)) positive and non zero, and the foregoing inequality is equivalent to

$$\mu_{2p+2} / \mu_{2p} \geqslant \mu_{2p} / \mu_{2p-2} .$$

Now the ratio of successive terms of the series

$$\lambda^2 \mu_2 + \lambda^4 \mu_4 + \lambda^6 \mu_6 + \cdots \cdot$$

is $(\mu_{2p+2} / \mu_p) \lambda^2$, which is therefore $\geqslant (\mu_4 / \mu_2) \lambda^2$.

If we choose $\lambda = \sqrt{\mu_2 / \mu_4}$ the terms of this series do not decrease, so that the series cannot converge. As this is a contradiction of our previous conclusion, it follows that $D(\lambda)$ has at least one root, of absolute magnitude $\leq \sqrt{\mu_2 / \mu_4}$.

2.5.2 Orthogonal Property of Characteristic Functions:

Let $u_1(x')$ be a characteristic function corresponding to a root λ_1, and let $u_2(x')$ be a characteristic function corresponding to a different root λ_2. Then

$$\int_a^b u_1(x')\, u_2(x')\, dx' = 0 .$$

For, by definition

$$u_1(x') = \lambda_1 \int_a^b K(x' \cdot \xi')\, u_1(\xi')\, d\xi'$$

$$u_2(x') = \lambda_2 \int_a^b K(x' \cdot \xi')\, u_2(\xi')\, d\xi'$$

Multiply the first of these equations by $\lambda_2 u_2(x')$, and the second by $\lambda_1 u_1(x')$. Subtract, and integrate with respect to the n variables x' between a and b. There results

$$(\lambda_2-\lambda_1)\int_a^b u_1(x')\,u_2(x')\,dx'$$

$$=\lambda_1\lambda_2\int_a^b\int_a^b\left\{K(x'\cdot\xi')u_1(\xi')u_2(x')-K(x'\cdot\xi')u_2(\xi')u_1(x')\right\}dx'd\xi'$$

$$=\lambda_1\lambda_2\int_a^b\int_a^b\left\{K(x'\cdot\xi')u_1(\xi')u_2(x')-K(\xi'\cdot x')u_1(x')u_2(\xi')\right\}dx'd\xi'$$

(since $K(x'\cdot\xi')=K(\xi'\cdot x')$)

$$=0$$

Since $\lambda_2\neq\lambda_1$,

$$\int_a^b u_1(x')\,u_2(x')\,dx'=0.$$

<u>2.5.3</u> <u>Theorem:</u> The characteristic numbers of a symmetrical kernel are all real.

If $D(\lambda)=0$ has a complex root $\mu+i\nu$, then (the coefficients in the power series for $D(\lambda)$ being real) $\mu-i\nu$ is also a root. Let $v(x')+iw(x')$ be a characteristic function corresponding to the root $\mu+i\nu$. i.e.

$$v(x')+iw(x')=(\mu+i\nu)\int_a^b K(x'\cdot\xi')\left\{v(\xi')+iw(\xi')\right\}d\xi'$$

Separating real and imaginary parts, and recombining, we see that $v(x')-iw(x')$ is a characteristic function corresponding to the root $\mu-i\nu$. Applying the result of 2.5.2, we have

$$\int_a^b\left\{v^2(x')+w^2(x')\right\}dx'=0.$$

This is impossible, since it implies that v and w both vanish identically. (If they did so, then $v+iw$ could not be a characteristic function, which by definition is non-zero)

2.5.4 Theorem: Every characteristic number of a symmetric kernel has a finite index.

Before proving this theorem - which we have already encountered in 2.3 - we require to derive the following

Lemma (Bessel's Inequality). Let $u_1(x'), u_2(x'), \cdots u_p(x')$ be orthogonal and normal in the ranges $a \leqslant x' \leqslant b$. i.e., let

$$\left.\begin{aligned}\int_a^b u_i(x')\,u_j(x')\,dx' &= 0 \quad \text{if } i \neq j \\ &= 1 \quad \text{if } i = j\end{aligned}\right\} \; [i,j = 1,2,\cdots p]$$

Suppose we wish to find the best representation (in the sense of "least squares") of a function $f(x')$ by the finite series $F(x') = c_1 u_1(x') + c_2 u_2(x') + \cdots + c_p u_p(x')$, where the c's are constants. We require, that is, to minimise $J = \int_a^b \{f(x') - F(x')\}^2 dx'$.

Differentiating, and using the orthogonal and normal properties $\partial J/\partial c_i = 2c_i - 2\int_a^b f(x')\,u_i(x')\,dx'$

Putting $\partial J/\partial c_i = 0$, we determine our constants as

$$c_i = \int_a^b f(x')\,u_i(x')\,dx' \quad (i = 1,2,\cdots p)$$

Since $\partial^2 J/\partial c_i^2 = 2$; $\partial^2 J/\partial c_i \partial c_j = 0 \; (i \neq j)$ these constants make J a minimum, of value

$$J_{min} = \int_a^b \Big\{f(x') - \sum_{i=1}^{p} u_i(x') \int_a^b f(\xi')\,u_i(\xi')\,d\xi'\Big\}^2 dx'.$$

Expanding, and again utilising the orthogonal and normal relations, we obtain

$$J_{min} = \int_a^b \{f(x')\}^2 dx' - \sum_{i=1}^{p} \Big\{\int_a^b f(x')\,u_i(x')\,dx'\Big\}^2$$

Since $J_{min} \geqslant 0$, there results Bessel's Inequality

$$\sum_{i=1}^{p}\left\{\int_a^b f(x')\,u_i(x')\,dx'\right\}^2 \leq \int_a^b \{f(x')\}^2\,dx'.$$

Returning to our theorem, we note that if $K(x'\cdot y')$ is symmetrical, the characteristic numbers are all real. Corresponding to a particular number λ, let there be p real linearly independent characteristic functions. We can form from these p linear combinations (also characteristic functions, since they satisfy the same integral equation) which are orthogonal for the ranges $a \leq x' \leq b$. (Whittaker and Watson, fourth edition, p. 224). After normalising, denote these by

$$u_1(x'),\quad u_2(x'),\ \cdots\ \cdots\quad u_p(x')$$

Taking these as the u's of Bessel's Inequality, and taking $K(y'\cdot x')$ for $f(x')$ we obtain

$$\sum_{i=1}^{p}\left\{\int_a^b K(x'\cdot y')\,u_i(y')\,dy'\right\}^2 \leq \int_a^b \{K(x'\cdot y')\}^2\,dy'$$

But

$$u_i(x') = \lambda\int_a^b K(x'\cdot y')\,u_i(y')\,dy' \qquad [i = 1, 2, \cdots p]$$

so

$$\sum_{i=1}^{p}\{u_i(x')\}^2/\lambda^2 \leq \int_a^b \{K(x'\cdot y')\}^2\,dy'.$$

Integrate with respect to the n variables x' between a and b. Since the u's are normal over these ranges, we obtain

$$p/\lambda^2 \leq \int_a^b\int_a^b \{K(x'\cdot y')\}^2\,dx'dy'$$

or

$$p \leq \frac{1}{\lambda^2}\int_a^b\int_a^b \{K(x'\cdot y')\}^2\,dx'\,dy'.$$

Thus there is an upper limit to p, the number of linearly independent real characteristic functions corresponding to the root λ. Also, it is readily seen that the number of linearly independent complex characteristic functions cannot exceed the number of linearly independent real characteristic functions. Therefore the index of the root λ is finite.

2.5.5 Suppose we write down all the characteristic numbers (of a given symmetric kernel) whose index is unity. Let us add the other characteristic numbers, whose index exceeds one, according to the following rule: numbers with index k are written down k times $(k = 2, 3, \cdots)$. We thus form a set $\{\lambda\}$ or

$$\lambda_1, \lambda_2, \cdots\cdots \lambda_i \cdots\cdots$$

of characteristic numbers; not necessarily all distinct. The number of occasions on which a given value appears is the index of that characteristic number. Since the index is always finite, we can write down the complete set of characteristic numbers.

Let us also write down the set of characteristic functions corresponding respectively to the characteristic numbers $\lambda_1, \lambda_2, \cdots$ (When we have a characteristic number of index k, we form k linear combinations, of the corresponding characteristic functions, which are mutually orthogonal in the ranges $a \leq x' \leq b$, as in Section 2.5.4) This set, like the set $\{\lambda\}$ can be written down completely; and every pair of members is orthogonal over the ranges $a \leq x' \leq b$; by (2.5.2). A characteristic function remains a characteristic function, of course, if it is multiplied by any numerical constant; we may therefore introduce suitable numerical factors so that the members of the set are normal for the ranges $a \leq x' \leq b$. Denote the normalised set by $\{u(x')\}$ or

$$u_1(x'),\ u_2(x'),\ \text{------}\ u_i(x'),\ \text{-----}$$

Let us recapitulate the properties of our two sets.

(i) They are complete.

(ii) Each member of $\{\lambda\}$ is a root of $D(\lambda) = 0$.

(iii) No member of $\{u(x')\}$ is identically zero.

(iv) A member λ_i of $\{\lambda\}$ is connected with the corresponding member $u_i(x')$ of $\{u(x')\}$ by the equation

$$u_i(x') = \lambda_i \int_a^b K(x' \cdot \xi')\, u_i(\xi')\, d\xi'$$

where K is symmetrical, in the sense that $K(x' \cdot \xi') = K(\xi' \cdot x')$.

(v) $$\left.\begin{aligned} \int_a^b u_i(x')\, u_j(x')\, dx' &= 0 \quad \text{if } i \neq j \\ &= 1 \quad \text{if } i = j \end{aligned}\right\}$$

2.6 The Development of A Symmetric Kernel: Let $K(x' \cdot y') = K(y' \cdot x')$. Form the sets $\{\lambda\}$ and $\{u(x')\}$ as described above. We shall show that

$$K(x' \cdot y') = \frac{u_1(x')\, u_1(y')}{\lambda_1} + \frac{u_2(x')\, u_2(y')}{\lambda_2} + \text{------}$$

provided that the series on the right is uniformly convergent everywhere in the ranges $a \leq x' \leq b$, $a \leq y' \leq b$.

Suppose this condition is satisfied. Since the functions u are continuous, the function represented by the series is continuous in all its variables. It is also symmetric. So

$$Q(x' \cdot y') \equiv K(x' \cdot y') - \sum_i \frac{u_i(x')\, u_i(y')}{\lambda_i} \tag{8}$$

is symmetric, continuous and finite everywhere in the ranges $a \leq x' \leq b$, $a \leq y' \leq b$. Assume that the theorem is false, and that, consequently, $Q(x' \cdot y')$ is not identically zero. By Schmidt's theorem (2.5.1) it has at least characteristic number, say c. Let the corresponding characteristic function be $\psi(x')$ i.e.,

$$\psi(y') = c\int_a^b Q(y' \cdot z')\, \psi(z')\, dz' \tag{9}$$

Multiply (8) by $u_i(y')$ and integrate with respect to the n variables y' between a and b. We obtain

$$\int_a^b Q(x' \cdot y')\, u_i(y')\, dy' = \int_a^b K(x' \cdot y')\, u_i(y')\, dy' - \frac{u_i(x')}{\lambda_i}$$

(using the orthogonal and normal properties of the u's)

$$= \frac{u_i(x')}{\lambda_i} - \frac{u_i(x')}{\lambda_i} = 0 \tag{10}$$

Again, from (9), we have

$$\begin{aligned}\int_a^b \psi(y')\, u_i(y')\, dy' &= c\int_a^b u_i(y')\, dy' \int_a^b Q(y' \cdot z')\, \psi(z')\, dz' \\ &= c\int_a^b \psi(z')\, dz' \int_a^b Q(y' \cdot z')\, u_i(y')\, dy' \\ &= 0\end{aligned} \tag{11}$$

(using (10) and recalling that Q is symmetric)

Finally, multiply (8) by $\psi(y')$ and integrate; whence

$$\int_a^b Q(x' \cdot y')\, \psi(y')\, dy' = \int_a^b K(x' \cdot y')\, \psi(y')\, dy'$$

or, from (9)

$$\frac{\psi(x')}{c} = \int_a^b K(x' \cdot y')\, \psi(y')\, dy'$$

Now ψ is continuous and not identically zero (since it was defined as a characteristic function) so that this last equation shows that it is a characteristic function of $K(x' \cdot y')$. It must therefore be linearly dependent on a finite number of the u's. But this is impossible since ψ is orthogonal to all the u's (equation(11)). Hence $Q \equiv 0$ and the theorem is proved.

2.6.1 Generalisation of the Preceding Theorem: The theorem just proved, while valuable, is rather restricted since it demands the uniform convergence of $\sum_i u_i(x')u_i(y')/\lambda_i$. The following theorem is somewhat more general:

If the series $\sum_i u_i(x')\,u_i(y')/\lambda_i$ converges uniformly in Cesaro's sense, with index r, then its sum (Cr) is $K(x' \cdot y')$

If the series converges uniformly (Cr) to the sum S then, by definition,

$$\lim_{k\to\infty} \sum_{i=1}^{k} \frac{u_i(x')\,u_i(y')}{\lambda_i\, b_{ik}}$$

converges uniformly to the sum S, where

$$b_{ik} = \left(1+\frac{r}{k+1-i}\right)\left(1+\frac{r}{k+2-i}\right)\left(1+\frac{r}{k+3-i}\right)\cdots\left(1+\frac{r}{k-1}\right)$$

Write

$$Q_k(x' \cdot y') \equiv K(x' \cdot y') - \sum_{i=1}^{k} \frac{u_i(x')\,u_i(y')}{\lambda_i\, b_{ik}} \qquad (12)$$

and let

$$Q(x' \cdot y') = \lim_{k\to\infty} Q_k(x' \cdot y')$$

By definition, $Q(x' \cdot y')$ converges uniformly in the ranges $a \leq x' \leq b$, $a \leq y' \leq b$, and each term in the series representing Q is continuous. Therefore Q itself is continuous, and also $Q(x' \cdot y') = Q(y' \cdot x')$, since K is symmetric.

Suppose that the theorem is false, and that $Q \not\equiv 0$. By Schmidt's theorem it has at least one characteristic number c, and a corresponding characteristic function $\psi(x')$ i.e.,

$$\psi(x') = c\int_a^b Q(x' \cdot y')\,\psi(y')\,dy' \qquad (13)$$

Multiplying (12) by $u_j(y')$ and integrating with respect to the n variables y' between a and b, we have

$$\int_a^b Q_k(x' \cdot y')u_j(y')\,dy' = \int_a^b K(x' \cdot y')u_j(y')\,dy' - \sum_{i=1}^{k} \frac{u_i(x')}{\lambda_i b_{ik}} \int_a^b u_i(y')u_j(y')\,dy'$$

$$= \frac{u_j(x')}{\lambda_j} - \frac{u_j(x')}{\lambda_j b_{jk}} \qquad (j \leq k)$$

Since $\sum_i u_i(x')\,u_i(y') / \lambda_i b_{ik}$ is uniformly convergent, this result (depending as it does on a term by term integration) remains valid as $k \to \infty$, when we have

$$\int_a^b Q(x' \cdot y')u_j(y')\,dy' = \lim_{k\to\infty}\left(\frac{u_j(x')}{\lambda_j} \cdot \frac{b_{jk}-1}{b_{jk}}\right) \qquad (14)$$

If we multiply (13) by $u_j(x')$ and integrate, there results

$$\int_a^b \psi(x')u_j(x')\,dx' = c\int_a^b u_j(x')\,dx' \int_a^b Q(x' \cdot y')\,\psi(y')\,dy'$$

$$= c\int_a^b \psi(y')\,dy' \int_a^b Q(x' \cdot y')\,u_j(x')\,dx'$$

$$= c\int_a^b \psi(y')\,dy' \left\{\lim_{k\to\infty} \frac{u_j(y')}{\lambda_j}\,\frac{b_{jk}-1}{b_{jk}}\right\}$$

(on using (14) and recalling the symmetry of Q)

$$= \frac{c}{\lambda_j}\left(\lim_{k\to\infty}\frac{b_{jk}-1}{b_{jk}}\right)\int_a^b \psi(y')\,u_j(y')\,dy'$$

So either

$$\text{(A)} \quad \frac{c}{\lambda_j} \lim_{k\to\infty} \frac{b_{jk-1}}{b_{jk}} = 1$$

or $$\text{(B)} \quad \int_a^b \psi(y')\, u_j(y')\, dy' = 0 .$$

Suppose (A) holds. Then Q has as many characteristic numbers as has K (previously we could only say that it had at least one) and they are given by

$$c = \lambda_j \lim_{k\to\infty} \frac{b_{jk}}{b_{jk-1}}$$

Now from the definition of the b's, we obtain

$$\frac{b_{jk}}{b_{jk-1}} = \frac{(k+1-j+r)(k+2-j+r)\cdots(k-1+r)}{(k+1-j+r)(k+2-j+r)\cdots(k-1+r)-(k+1-j)(k+2-j)\cdots(k-1)}$$

Multiplying out in powers of r,

$$\frac{b_{jk}}{b_{jk-1}} = \frac{O(k^{j-1}) + r\,O(k^{j-2}) + \cdots + r^{j-1}}{r\,O(k^{j-2}) + \cdots + r^{j-1}} = O(k)$$

as $k \to \infty$.

Hence $|c| = O(k)$ as $k \to \infty$. i.e., <u>all</u> the characteristic numbers of Q are indefinitely large. But this is impossible, since at least one root is less in absolute magnitude than $\sqrt{U_2/U_4}$, where U_2 and U_4 are defined in terms of the iterated functions of Q - see 2.5.1. We must therefore reject hypothesis (A) and consider (B); viz.,

$$\int_a^b \psi(y')\, u_j(y')\, dy' = 0 \qquad (15)$$

Let us multiply (12) by $\psi(y')$ and integrate. We obtain

$$\int_a^b Q_k(x'.y')\psi(y')dy' = \int_a^b K(x'.y')\psi(y')dy' - \sum_{i=1}^{k} \frac{u_i(x')}{\lambda_i b_{ik}} \int_a^b u_i(y')\psi(y')dy'$$

$$= \int_a^b K(x'.y')\,\psi(y')\,dy' \qquad \text{from (15)}$$

Proceeding to the limit, $k \to \infty$, and remembering (13), it follows that

$$\frac{\psi(x')}{c} = \int_a^b K(x' \cdot y')\, \psi(y')\, dy'$$

As in the preceding section, this equation is incompatible with (15), and we conclude that $Q \equiv 0$ i.e., that the sum (C r) of $\sum_i u_i(x') u_i(y')/\lambda_i$ is $K(x' \cdot y')$

This theorem includes the preceding one, since Cesaro uniform convergence with r = 0 is 'ordinary' uniform convergence

<u>2.7</u> <u>Theorem</u>: The set $\{u(x')\}$ is closed.

To prove this theorem, we require to show that there is no continuous non-zero function $h_0(x')$ other than members of the set $\{u(x')\}$ for which

$$\int_a^b h_0(x')\, u_i(x')\, dx' = 0 \quad (i = 1, 2, \cdots) \qquad (16)$$

Suppose that $\{u(x')\}$ is not a closed set, and that a non-zero continuous function h_0 does exist which satisfies (16). Then $h(x') = A\, h_0(x')$ also satisfies (16), where A is any numerical constant.

Consider

$$Q(x' \cdot y') = K(x' \cdot y') - h_0(x')\, h_0(y')$$

Since K is symmetric and continuous, and since h_0 is assumed continuous, Q is symmetric and continuous. It has therefore at least one characteristic number c, with corresponding characteristic function $\psi(x')$ i.e.,

$$\psi(x') = c\int_a^b Q(x' \cdot y')\, \psi(y')\, dy'$$

Multiply by $h(x') = A\,h_0(x')$ and integrate. We obtain

$$\int_a^b \psi(x')h(x')dx' = c\int_a^b h(x')dx'\int_a^b Q(x'\cdot y')\psi(y')dy'$$
$$= c\int_a^b \psi(y')dy'\int_a^b \{K(x'\cdot y') - h_0(x')h_0(y')\}h(x')dx'$$
$$= c\int_a^b \psi(y')dy'\int_a^b \{K(y'\cdot x')h(x') - h_0(x')h(x')h_0(y')\}dx' \quad (17)$$

Suppose that the series

$$\sum_i \frac{u_i(x')u_i(y')}{\lambda_i}$$

is uniformly convergent (C r) in the ranges $a \le x' \le b$, $a \le y' \le b$.

From the last section we know that

$$K(y'\cdot x') = \lim_{k\to\infty} \sum_{i=1}^{k} \frac{u_i(x')u_i(y')}{\lambda_i b_{ik}},$$

the series on the right being uniformly convergent throughout $a \le x' \le b$, $a \le y' \le b$. Hence, integrating term by term

$$\int_a^b K(y'\cdot x')h(x')dx' = 0.$$

(by equation (16)) and (17) reduces to

$$\int_a^b \psi(x')h(x')dx' = -c\left(\int_a^b \psi(y')h(y')dy'\right)\left(\int_a^b h_0(x')h(x')dx'\right)$$

Therefore, either

(A) $\int_a^b h_0(x')h(x')dx' = -1/c.$

or (B) $\int_a^b \psi(x')h(x')dx' = 0$

Suppose hypothesis (A) is true. This implies that

$$\int_a^b h_0(x')h(x')dx' = A\left\{\int_a^b h_0^2(x')dx'\right\}$$

can only assume certain particular values, viz., the values $-1/c$ where c is a characteristic number of Q. But this cannot be, since A is <u>any</u> number and $h_0 \not\equiv 0$ by hypothesis. We therefore reject hypothesis (A) and consider (B), that is $\int_a^b \psi(y')\,h(y')\,dy' = 0$.

Substituting this result in the equation

$$\psi(x') = c\int_a^b Q(x'\cdot y')\,\psi(y')\,dy'$$
$$= c\int_a^b \left\{K(x'\cdot y') - h_0(x')\,h_0(y')\right\}\psi(y')\,dy'$$

we obtain

$$\psi(x') = c\int_a^b K(x'\cdot y')\,\psi(y')\,dy' \quad .$$

Thus c is a characteristic number of K. It follows that <u>all</u> the characteristic numbers of K and of Q are respectively equal. Now these numbers are, respectively, the roots of

$$D_K(\lambda) = 1 - \lambda\int_a^b K(\xi_1'\cdot\xi_1')\,d\xi_1' + \frac{\lambda^2}{2!}\int_a^b\int_a^b K\begin{pmatrix}\xi_1' & \xi_2' \\ \xi_1' & \xi_2'\end{pmatrix} d\xi_1'\,d\xi_2' - \cdots$$
$$= 0$$

and of

$$D_Q(\lambda) = 1 - \lambda\int_a^b Q(\xi_1'\cdot\xi_1')\,d\xi_1' + \frac{\lambda^2}{2!}\int_a^b\int_a^b Q\begin{pmatrix}\xi_1' \cdot \xi_2' \\ \xi_1' \; \xi_2'\end{pmatrix} d\xi_1'\,d\xi_2' - \cdots$$
$$= 0$$

We require that these equations have identical roots. As the left hand side of each begins with unity, the coefficients of $\lambda, \lambda^2, \lambda^3, \cdots\cdots$ in the one must equal, respectively, the coefficients of $\lambda, \lambda^2, \lambda^3, \cdots$ in the other. To achieve this, it is obviously sufficient that $K \equiv Q$. This condition is also necessary, since (equating the coefficients of λ)

$$\int_a^b K(\xi_1'\cdot\xi_1')\,d\xi_1' = \int_a^b Q(\xi_1'\cdot\xi_1')\,d\xi_1'$$
$$= \int_a^b \left\{K(\xi_1'\cdot\xi_1') - h_0^2(\xi_1')\right\} d\xi_1'$$

whence $\int_a^b h_0^2(\xi_1')\,d\xi_1' = 0.$

As h_0 was assumed continuous, we must have $h_0 \equiv 0$ and, therefore, $K \equiv Q$.

It follows that the set $\{u(x')\}$ is closed.

2.8 Expansion of An Arbitrary Function in Terms of $\{u(x')\}$: Let us try to express an arbitrary function $f(x')$ in terms of the members of the closed set $\{u(x')\}$ in the form, say,

$$f(x') = \sum_i c_i u_i(x')$$

where the coefficients c_i are constants. If we multiply by $u_j(x')$ and integrate with respect to the n variables x' between a and b, we obtain at once the value of these constants, viz.,

$$c_j = \int_a^b f(x') u_j(x') dx' \qquad (j = 1, 2, \cdots)$$

since the set $\{u(x')\}$ is orthogonal and normal. The constants c_j $(j = 1, 2, \cdots)$ are termed the "generalised Fourier coefficients" of $f(x')$ relative to the set $\{u(x')\}$.

It is a well known result that the foregoing expansion exists provided $f(x')$ is a function of class L^2 over the ranges $a \leq x' \leq b$ i.e. provided

$$\int_a^b |f(x')|^2 dx' \text{ is finite.}$$

($f(x')$ may be discontinuous, and, if necessary, the integral may be interpreted in Lebesgue's sense). If, in particular, $f(x')$ is continuous and finite, and if a and b are finite, then the condition is obviously fulfilled.

The converse question - given a set of constants $\{c\}$ or $c_1, c_2, \ldots\ldots c_i \ldots\ldots$ does there exist a function of which they are the generalised Fourier coefficients relative to $\{u(x')\}$? - has been answered by Riesz (Göttingen Nachrichten, 1907) in the following

Theorem: Given the set of constants $\{c\}$, the necessary and sufficient condition that a function $f(x')$ exist, of the class L^2 over the ranges $a \leq x' \leq b$, whose generalised Fourier coefficients relative to $\{u(x')\}$ are $\{c\}$ is that

$$\sum_i c_i^2$$

be convergent.

2.9 The Homogeneous Equation $\int_a^b g(y') K(x' y') dy' = f(x')$ with Symmetrical Kernel: This equation, in which K and f are given, and in which g is unknown, frequently arises in statistical theory. The necessary and sufficient condition for a solution to exist, and the expression for this solution, have been obtained by Picard (Comptes-Rendus de l'Académie des Sciences de Paris, Vol. 148, p. 1563 (1909)). His results are as follows:

Theorem: The necessary and sufficient condition that the integral equation with symmetrical kernel

$$\int_a^b g(y') K(x' y') dy' = f(x') \tag{18}$$

should have a solution of the class L^2 is that

$$\sum_i \lambda_i^2 c_i^2$$

should converge, where $\{\lambda\}$ is the set of characteristic numbers of K, and $\{c\}$ is the set of generalised Fourier coefficients of $f(x')$ relative to $\{u(x')\}$.

When this condition is fulfilled, a solution of (18) is

$$g(y') = \sum_i \lambda_i c_i u_i(y') \tag{19}$$

and this is the only solution of class L^2.

It is easy to demonstrate that, at least formally, (19) satisfies the integral equation (18). Assume that a solution exists of the form

$$g(y') = \sum_i a_i u_i(y')$$

where the a_i's are constants. We also have

$$f(x') = \sum_i c_i u_i(x')$$

$$K(x', y') = \sum_i \frac{u_i(x') u_i(y')}{\lambda_i}$$

and we may suppose for the moment that this last expression is uniformly convergent when $a \leq x' \leq b$, $a \leq y' \leq b$. If we substitute these three expressions in (18) we obtain

$$\int_a^b \left\{ \sum_i a_i u_i(y') \right\} \left\{ \sum_j \frac{u_j(x') u_j(y')}{\lambda_j} \right\} dy' = \sum_k c_k u_k(x').$$

Since the set $\{u(x')\}$ is orthogonal and normal, this reduces (at least formally) to

$$\sum_i \frac{a_i u_i(x')}{\lambda_i} = \sum_k c_k u_k(x')$$

Multiplying by $u_j(x')$ and integrating

$$a_j / \lambda_j = c_j, \quad \text{or} \quad a_j = \lambda_j c_j \quad (j = 1, 2, \cdots)$$

Therefore $g(y') = \sum_j \lambda_j c_j u_j(y')$ satisfies (18) formally.

2.9.1 <u>Necessity of Picard's Condition:</u> Let $g(y')$ be a solution of class L^2, so that

$$f(x') = \int_a^b K(x'.y')\, g(y')\, dy'$$

Since $f(x') = \sum_i c_i u_i(x')$ we have

$$c_i = \int_a^b f(x')\, u_i(x')\, dx'.$$

Substituting the expression for $f(x')$ into this last equation, we obtain

$$c_i = \int_a^b u_i(x')\, dx' \int_a^b K(x'.y')\, g(y')\, dy'$$
$$= \int_a^b g(y')\, dy' \int_a^b K(y'.x')\, u_i(x')\, dx'$$

since $K(x'.y') = K(y'.x')$.

Now the set $\{u(x')\}$ is the set of characteristic functions of $K(x'.y')$. Therefore

$$u_i(y') = \lambda_i \int_a^b K(y'.x')\, u_i(x')\, dx'$$

whence

$$c_i = \frac{1}{\lambda_i} \int_a^b g(y')\, u_i(y')\, dy'$$

Thus $\{\lambda c\}$ is the set of generalised Fourier coefficients, relative to $\{u(x')\}$, of a function $g(y')$ of class L^2. Therefore, by Riesz' Theorem (Section 2.8)

$$\sum_i \lambda_i^2 c_i^2$$

is convergent.

2.9.2 <u>Sufficiency of Picard's Condition:</u> Let

$$\sum_i \lambda_i^2 c_i^2$$

converge. By Riesz' theorem, $\{\lambda c\}$ is the set of generalised Fourier coefficients, relative to $\{u(x')\}$, of some function of class L^2, say of $g(y')$. Write

$$F(x') = \int_a^b K(x'.y')\, g(y')\, dy'$$

We shall find the generalised Fourier coefficients of $F(x')$. If these are $\{C\}$ then

$$\begin{aligned} C_i &= \int_a^b F(x')\, u_i(x')\, dx' \\ &= \int_a^b u_i(x')\, dx' \int_a^b K(x'.y')\, g(y')\, dy' \\ &= \int_a^b g(y')\, dy' \int_a^b K(x'.y')\, u_i(x')\, dx' \end{aligned}$$

Now

$$K(x'.y') = K(y'.x') = \sum_j \frac{u_j(x')\, u_j(y')}{\lambda_j}$$

(supposing this series is uniformly convergent). By the orthogonal and normal properties, therefore

$$\int_a^b K(x'.y')\, u_i(x')\, dx' = u_i(y')/\lambda_i$$

whence

$$C_i = \frac{1}{\lambda_i} \int_a^b g(y')\, u_i(y')\, dy'$$

But, by definition of the function g,

$$\int_a^b g(y')\, u_i(y')\, dy' = \lambda_i c_i$$

So

$$C_i = c_i$$

i.e., the generalised Fourier coefficients, relative to $\{u(x')\}$, of the function $F(x')$ are identical with those of $f(x')$. Therefore, since the set $\{u(x')\}$ is closed (Section 2.7)

$$F(x') = f(x')$$

or

$$f(x') = \int_a^b K(x'.y')\, g(y')\, dy'$$

i.e., when Picard's condition holds, $\sum_i \lambda_i c_i u_i(x')$ is a valid solution, of the class L^2, of the integral equation (18). It is clear, too, that this is the only solution of class L^2.

If the series

$$\sum_i u_i(x')u_i(y')/\lambda_i$$

is not uniformly convergent, but merely uniformly convergent (C r) [+70] we may repeat the foregoing analysis step by step; using, where appropriate, the uniformly convergent series

$$\lim_{k\to\infty} \sum_{i=1}^{k} \frac{u_i(x')\,u_i(y')}{\lambda_i b_{ik}}.$$

We find that the equation (18) is satisfied formally by

$$g(y') = \sum_i a_i u_i(y')$$

where

$$a_i = \lim_{k\to\infty} \lambda_i c_i b_{ik} = \lambda_i c_i \lim_{k\to\infty} b_{ik} = \lambda_i c_i$$

since

$$\lim_{k\to\infty} b_{ik} = \lim_{k\to\infty}\left(1+\frac{\tau}{k+1-i}\right)\left(1+\frac{\tau}{k+2-i}\right)\cdots\left(1+\frac{\tau}{k-1}\right) = 1$$

We thus have the same solution as in the simple case treated already. Continuing the analysis, we readily find that this solution is valid, is of class L^2, and is the only solution of class L^2, provided $\sum_i \left(\lim_{k\to\infty} \lambda_i^2 c_i^2 b_{ik}^2\right)$ converges; this condition being both necessary and sufficient.

<u>2.9.3</u> <u>Corollary:</u> In the foregoing, let $f(x') \equiv 0$. Then

$$c_i = 0 \quad [i = 1, 2, \cdots]$$

and $\sum \lambda_i^2 c_i^2$ is convergent. Hence the equation

$$\int_a^b g(y')\,K(x';y')\,dy' = 0 \tag{20}$$

has only one solution of class L^2, viz.,

$$g(y') = \sum_i \lambda_i c_i u_i(y') \equiv 0.$$

Since a continuous, bounded function is of class L^2 over any finite range of the variables, it follows that the only continuous solution of the integral equation (2) is

$$g \equiv 0.$$

2.10 The Homogeneous Equation $\int_a^b g(y')H(x'y')dy' = h(x')$ With Unsymmetrical Kernel: The integral equation

$$\int_a^b g(y')H(x'y')dy' = h(x'), \qquad (21)$$

in which the kernel H is unsymmetrical in x' and y', can be solved by solving an "associated" equation with a symmetrical kernel, viz.,

$$\int_a^b g(y')K(x'y')dy' = f(x') \qquad (22)$$

where $K(x'y') = \int_a^b H(\xi'x')H(\xi'y')d\xi' = K(y'x')$

and $f(x') = \int_a^b h(\xi')H(\xi'x')d\xi'$.

Theorem: The solutions of (21) and of (22) are identical. Suppose that a solution $g(y')$ of equation (21) exists. Multiplying this equation by $H(x'\xi')$ and integrating, we find

$$\int_a^b H(x'\xi')dx'\int_a^b g(y')H(x'y')dy' = \int_a^b h(x')H(x'\xi')dx'$$

or

$$\int_a^b g(y')K(\xi'y')dy' = f(\xi').$$

Hence any solution of (21) is also a solution of (22), and the latter can be solved by the method of Section 2.9.

We shall now demonstrate the converse result - that the solution of class L^2, of (22) satisfies (21).

Denote the former by $g_1(y')$ so that

$$\int_a^b g_1(y')\,K(x'\cdot y')\,dy' = f(x')$$

or (by definition of K and of f)

$$\int_a^b g_1(y')\,dy' \int_a^b H(\xi'\cdot x')\,H(\xi'\cdot y')\,d\xi' = \int_a^b h(\xi')\,H(\xi'\cdot x')\,d\xi'$$

Hence

$$\int_a^b H(\xi'\cdot x')\,d\xi' \int_a^b g_1(y')\,H(\xi'\cdot y')\,dy' = \int_a^b h(\xi')\,H(\xi'\cdot x')\,d\xi'$$

or

$$\int_a^b H(\xi'\cdot x')\,d\xi' \left\{ \int_a^b g_1(y')\,H(\xi'\cdot y')\,dy' - h(\xi') \right\} = 0$$

Put

$$j(\xi') = \int_a^b g_1(y')\,H(\xi'\cdot y')\,dy' - h(\xi')$$

so that

$$\int_a^b H(\xi'\cdot x')\,j(\xi')\,d\xi' = 0 \tag{23}$$

Multiply by $H(\eta'\cdot x')$ and integrate. We have

$$\int_a^b H(\eta'\cdot x')\,dx' \int_a^b H(\xi'\cdot x')\,j(\xi')\,d\xi' = 0$$

or

$$\int_a^b j(\xi')\,d\xi' \int_a^b H(\xi'\cdot x')\,H(\eta'\cdot x')\,dx' = 0$$

or [writing $\int_a^b H(\xi'\cdot x')\,H(\eta'\cdot x')\,dx' = K_1(\xi'\cdot\eta') = K_1(\eta'\cdot\xi')$]

$$\int_a^b j(\xi')\,K_1(\xi'\cdot\eta')\,d\xi' = 0 \tag{24}$$

Now, we have seen that if (23) possesses a solution, this solution also satisfies (24). By Section 2.9.3, equation (24) (with a symmetric kernel) has only one solution of class L^2, viz., $j \equiv 0$.

By actual substitution, we see that this does satisfy (23). In other words

$$\int_a^b g_1(y')\,H(\xi'\cdot y')\,dy' = h(\xi') \quad ;$$

that is, the solution, of class L^2, of (22) also satisfies (21)

2.10.1 Corollary: If a continuous solution of (21) is known, it is the only continuous solution. For if there were a second, equation (22) would possess two continuous (hence class L^2) solutions, in contradiction of Section 2.9.

2.11 Transformation of Variables: A known solution g of an integral equation

$$\int_a^b g(y')\,K(x'\cdot y')\,dy' = f(x')$$

can be used to obtain the solution of another integral equation

$$\int_{(c)}^{(d)} g^x(t')\,K^x(x'\cdot t')\,dt' = f(x'),$$

derived from the first by a transformation of the variables.

Let the transformation be

$$y_i = u_i(t_1 \cdot t_2 \cdots t_n) \qquad [i = 1, 2, \cdots n]$$

or, in abbreviated notation

$$y' = u'(t')$$

Substituting in the first equation (in which g is regarded as already known), we have

$$\int_{(c)}^{(d)} g[u'(t')]\,K\big(x'\cdot u'(t')\big)\,J\,dt' = f(x')$$

where (c), (d) denote the limits of integration of the variables $t_1, \cdots t_n$ corresponding to the limits a, b of $x_1, \cdots\cdots x_n$, and where J is the Jacobian

$$\frac{\partial(y_1 \cdots y_n)}{\partial(t_1 \cdots t_n)}$$

If we now put

$$g[u'(t')] \cdot J = g^{x}(t')$$
$$K\left(x' \cdot u'(t')\right) = K^{x}(x' \cdot t')$$

the last equation may be rewritten as

$$\int_{(c)}^{(d)} g^{x}(t') K^{x}(x' \cdot t') dt' = f(x')$$

which we may regard as an integral equation in an unknown function g^x. But if the solution g to the original equation has been found, we can at once determine g^x, the solution of this new equation.

Conversely, if the solution g^x of the last equation were known, we could deduce the solution of the first, on applying the transformation inverse to $y' = u'(t')$.

2.12. Singular Kernels:- It was stipulated at the beginning of this chapter, that the kernel $K(x' \cdot y')$ of our integral equations was continuous and finite throughout the ranges $a \leq x' \leq b$, $a \leq y' \leq b$. This restriction has been implicit in all our results hitherto. Looking back, one appreciates that the corner stone, upon which the subsequent development was built, was the convergence of the "associated series" of Section 2.1. This observation points the way in which our restriction on K may be relaxed. If K has singularities, but if the associated series (defined as before) still have a meaning, and remain absolutely and/or uniformly convergent, we expect that all the preceding theorems will be valid. The question of the solution of an integral equation with such a kernel has been investigated by Hilbert, (Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen, 1924, chapter 6), whose results may be summed up as follows:

All the results of this chapter remain valid if the kernel $K(x' \cdot t')$ has a finite number of singularities, the order of which is less than 1/2.

($K(x' \cdot t')$ is said to have a singularity of order α at say, $x_i = \omega_i$ if $[K(x' \cdot t')]_{x_i = \omega_i}$ is infinite, and if there exists a number $\alpha > 0$ such that $(x_i - \omega_i)^{\beta} \cdot K(x' \cdot t')$ is finite and continuous in the neighbourhood of $x_i = \omega_i$ when $\beta \geqslant \alpha$ but is infinite at $x_i = \omega_i$ when $\beta < \alpha$.

2.13 Integral Equations with Infinite Limits of Integration:- Thus far, we have consistently assumed that the ranges of integration were finite (an assumption demanded, in the first instance, in our proofs of the convergence of the associated series of the kernel). The consideration of an equation with infinite limits in the integral does not, however, introduce any radically new feature, since we may apply a transformation of variables which will make the range finite.

Suppose, for instance, that the range of all $2n$ variables is $(0,\infty)$ i.e. $0 \leq x'$; $0 \leq \xi'$. If we apply the transformation

$$y = \frac{x'}{1+x'} \quad ; \quad t' = \frac{\xi'}{1+\xi'}$$

the new range of the variables is $0 \leq y' \leq 1$; $0 \leq t' \leq 1$. The transformed kernel may, however, have a singularity at $y'=1$ (i.e., when any of the n variables of integration assumes the value unity). If the singularities are of order α where $0 \leq \alpha < \frac{1}{2}$ then all the preceding results remain valid. For instance, an equation of the form $\int_0^1 g^*(y')K^*(x'y')dy' = f(x')$ will have a unique solution of class L^2, and this solution, by section 2.11, will also be the unique solution of class L^2 of the original homogeneous equation with limits $(0,\infty)$ in the integral.

If the range of integration is $(-\infty, 0)$ or $(-\infty,\infty)$ we proceed in exactly the same manner; if the kernel, after applying a transformation which makes the limits of integration finite, has singularities of order less than 1/2, then the equation can be solved by the methods discussed above.

<u>2.13.1</u> <u>Probability Distribution as Kernel</u>:- In statistical work, it frequently happens that the kernel of an integral equation is a probability distribution, say $\varphi(x|\theta)$, continuous in both x and θ. Let the range of x be $(0,\infty)$. By the total probability condition

$$\int_0^\infty \varphi(x|\theta)\,dx = 1 \qquad \text{for all values of } \theta.$$

Suppose we transform this distribution by the transformation $x_1 = x/(1+x)$. The probability distribution of x_1 is consequently $\varphi_1(x_1|\theta)$ where

$$\varphi_1(x_1|\theta)\,dx_1 = \varphi\left(\frac{x_1}{1-x_1}\Big|\theta\right)\frac{dx_1}{(1-x_1)^2}, \quad 0 \le x_1 \le 1.$$

and the total probability condition becomes

$$\left.\begin{aligned} &\int_0^1 \varphi_1(x_1|\theta)\,dx_1 = 1 \\ \text{or}\quad &\int_0^1 \varphi\left(\frac{x_1}{1-x_1}\Big|\theta\right)\frac{dx_1}{(1-x_1)^2} = 1 \end{aligned}\right\}\text{for all values of } \theta. \tag{25}$$

Since φ is continuous, positive everywhere, and always <1, it follows that φ_1 is continuous everywhere except possibly at $x_1 = 1$, and is always positive. From (25) therefore, it is necessary that

$$\int_X^1 \varphi_1(x_1|\theta)\,dx_1 < 1$$

where X is any positive number less than 1, which we may choose arbitrarily close to 1. Hence $\varphi_1(1|\theta)$ is finite, and $\varphi_1(x_1|\theta)$ has <u>no</u> singularities.

If, therefore, we encounter an integral equation whose kernel is a probability distribution of range $(0,\infty)$ we may apply the transformation $x_1 = x/(1+x)$ and obtain a <u>non-singular</u> kernel in the transformed equation, for which the theorems of this chapter are valid.

<u>2.13.2</u> <u>Examples</u>: We give below some examples of transformed probability distributions.

Example 1. The normal distribution with mean zero

$$\varphi = \sqrt{2/\pi}\ \sigma^{-1} \exp\left(-x^2/2\sigma^2\right) \qquad (x \geq 0)$$

transforms (when $x \to x_1/\overline{1-x_1}$) into

$$\varphi_1(x_1|\sigma) = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sigma} \frac{\exp\left\{-\frac{1}{2\sigma^2}\left(\frac{x_1}{1-x_1}\right)^2\right\}}{(1-x_1)^2} \qquad (0 \leq x_1 < 1)$$

and, let us assert, $\varphi_1(1|\sigma) = 0$.

This function φ_1 is bounded and continuous when $0 \leq x_1 \leq 1$. This is obviously true when $0 \leq x_1 < 1$. Also

$$\lim_{x_1 \to 1-0} \varphi_1(x|\sigma) = \sqrt{2/\pi} \cdot \sigma^{-1} \lim_{x_1 \to 1-0} \frac{\exp\left[-\frac{1}{2\sigma^2}\left(\frac{x_1}{1-x_1}\right)^2\right]}{(1-x_1)^2}$$

$$= \sqrt{2/\pi}\ \sigma^{-1} \lim_{\epsilon \to 0+0} \frac{1}{\epsilon^2} \exp\left\{\frac{1}{2\sigma^2}\left(\frac{1-\epsilon}{\epsilon}\right)^2\right\}$$

$$= \sqrt{2/\pi}\ \sigma^{-1} \lim_{\epsilon \to 0+0} \frac{1}{\epsilon^2} \left\{\frac{1}{1 + \left(\frac{1-\epsilon}{\epsilon}\right)^2 \frac{1}{2\sigma^2} + \frac{1}{2!}\left(\frac{1-\epsilon}{\epsilon}\right)^4 \frac{1}{4\sigma^4} + O(\epsilon^{-6})}\right\}$$

$$= \sqrt{2/\pi}\ \sigma^{-1} \lim_{\epsilon \to 0+0} \frac{1}{\epsilon^2 + (1-\epsilon)^2/2\sigma^2 + O(\epsilon^{-2})}$$

$$= 0$$

$$= \varphi_1(1|\sigma).$$

Hence $\varphi_1(x_1|\sigma)$ is continuous and bounded throughout $0 \leq x_1 \leq 1$.

Example 2. The exponential distribution

$$\varphi = a \exp(-x/a) \quad , \quad (x \geq 0)$$

transforms [when $x \to x_1/\overline{1-x_1}$] into

$$\varphi_1(x_1|a) = \frac{a}{(1-x_1)^2} \exp\left\{-\frac{x_1}{a(1-x_1)}\right\} \qquad (0 \leq x_1 < 1)$$

and, let us assert, $\varphi_1(1|a) = 0$.

This function is bounded and continuous when $0 \leq x_1 < 1$.

Also

$$\lim_{x_1 \to 1-0} \varphi_1(x_1|a) = \lim_{\epsilon \to 0+0} \frac{a}{\epsilon^2} \exp\left(-\frac{1-\epsilon}{a\epsilon}\right)$$

$$= a \exp(1/a) \lim_{\epsilon \to 0+0} \frac{1}{\epsilon^2} \frac{1}{\exp(1/a\epsilon)}$$

$$= a \exp(1/a) \lim_{\epsilon \to 0+0} \left[\left\{\epsilon^2 + \frac{\epsilon}{a} + \frac{1}{2a^2} + O(\epsilon^{-1})\right\}^{-1}\right]$$

$$= 0$$

$$= \varphi_1(1|a)$$

Hence $\varphi_1(x_1|a)$ is continuous and bounded throughout $0 \leq x_1 \leq 1$.

<u>Example 3.</u> The Cauchy distribution

$$\varphi = \frac{1}{\pi} \frac{1}{1+(x-m)^2} \qquad (-\infty < x < \infty)$$

transforms $\left[\begin{array}{ll} \text{when } x \to x_1/\overline{1-x_1}, & x \geq 0 \\ \phantom{\text{when }} x \to x_1/\overline{1+x_1}, & x \leq 0 \end{array}\right]$ into

$$\varphi_1(x_1|m) = \frac{1}{\pi} \cdot \frac{1}{1+\left(\frac{x_1}{1-x_1} - m\right)^2} \cdot \frac{1}{(1-x_1)^2} \qquad (0 \leq x_1 < 1)$$

$$\varphi_1(x_1|m) = \frac{1}{\pi} \cdot \frac{1}{1+\left(\frac{x_1}{1+x_1} - m\right)^2} \cdot \frac{1}{(1-x_1)^2} \qquad (0 \geq x_1 > 1)$$

and, let us assert, $\varphi_1(1|m) = \varphi_1(-1|m) = 1/\pi$.

This function is obviously finite and continuous when $-1 < x_1 < 1$. Also

$$\lim_{x_1 \to 1-0} \varphi_1(x_1|m) = \frac{1}{\pi} \lim_{x_1 \to 1-0} \frac{1}{(1-x_1)^2 + \{x_1 - m(1-x_1)\}^2} = \frac{1}{\pi} = \varphi_1(1|m)$$

Similarly $\lim_{x_1 \to -1+0} \varphi_1(x_1 | m) = \varphi_1(-1 | m)$

Hence φ_1 is bounded and continuous throughout the range $-1 \leq x_1 \leq 1$.

2.14 Recapitulation:- Let us summarise the steps by which we can solve a homogeneous integral equation of the form $\int_a^b g(y') K(x' \cdot y') dy' = f(x')$.

If the range (a, b) is infinite, we apply a transformation of the variables so that it becomes finite. We verify that the transformed kernel has only a finite number of singularities, the order of these being $< \frac{1}{2}$. (or, if the function K is a probability distribution, we know that the transformed kernel has no singularities). The solution of the original equation is known once that of the transformed equation is determined.

If the kernel (after the foregoing transformation has been applied if necessary) is unsymmetrical in the vectors x', y', we "symmetrise" it as in Section 2.10. The solutions (of class L^2) of these two equations are identical.

We form the associated series $D(\lambda)$ of the symmetrised kernel, and abstract all the roots of $D(\lambda) = 0$. (There will be one root at least, and all are real). The characteristic functions corresponding to each root λ can be found in terms of the other associated series of the kernel, $D(x'/y'; \lambda); D(x_1' \cdot x_2' / y_1' \cdot y_2'; \lambda)$ .... etc., as in Section 2.3. We observe the smallest integer ρ for which the series $D(x_1' \cdots x_p' | y_1' \cdots y_p' ; \lambda)$ is not identically zero (λ being a particular characteristic number). This value of ρ is the index of that particular characteristic number, and there are ρ linearly independent characteristic functions to be determined, as in 2.3. If ρ exceeds 1, these functions are to be formed into linear combinations which are orthogonal over the range we are interested in. These, along with the other characteristic functions corresponding to roots of index 1, are to be normalised by suitable numerical factors. We then have the complete and closed set $\{u(x')\}$ of characteristic functions and the complete set $\{\lambda\}$ of characteristic numbers. We verify that the series $\sum_i u_i(x') u_i(y') / \lambda_i$ is uniformly convergent $(C\,r)$ where $r \geqslant 0$. If so, the unique solution of class L^2 of our equation with the symmetrised kernel can now be obtained, as in Section 2.9.

Chapter Three

THE ESTIMATION OF A SINGLE PARAMETER BY MEANS OF AN UNBIASSED STATISTIC OF MINIMUM VARIANCE

3.0 The probability distribution of the population we are sampling will be denoted by $\varphi(x|\theta)$, a function continuous in both x and θ, and possessing continuous derivatives, with respect to θ, of all orders. The range of the variate x will be taken as independent of θ, though it may be either finite or infinite.

The probability that an observation of the variate will have a value within the small range $x_1 \pm \frac{1}{2}dx_1$ is

$$\varphi(x_1|\theta)\,dx_1$$

Since an observed value certainly lies somewhere within the total range of x, we have the "total probability condition"

$$\int \varphi(x|\theta)\,dx = 1 \quad \text{for all values of } \theta,$$

the integral extending over all possible values of x.

Suppose we take a sample of n random observations $x_1, x_2, \cdots x_n$ (which may be denoted by the vector x'). The probability that n observations will lie respectively within the small ranges $x_i \pm \frac{1}{2}dx_i\ [i = 1, 2, \cdots n]$ is, of course

$$\prod_{i=1}^{n} \varphi(x_i|\theta)\,dx_i = \Phi(x'|\theta)\,dx' \quad \text{say.}$$

On integrating, over their total range, with respect to the n variables, we obtain

$$\int \Phi(x'|\theta)\,dx' = \int \varphi(x_1|\theta)dx_1 \int \varphi(x_2|\theta)\,dx_2 \cdots\cdots \int \varphi(x_n|\theta)\,dx_n$$

$$= 1 \qquad \text{for all values of } \theta.$$

Were the value of θ known, we could calculate the probability of our n random observations lying within any given limits. The problem of estimation is in a sense the converse to this. The value of θ is quite unknown, and we must use the observations comprising the sample in order to infer it.

In development of the criteria of Chapter 1, we say (by way of a priori assertion) that the "best" value of θ is provided by a statistic $T(x')$ – i.e. a function of the observations - which has the following properties :

(a) It is a symmetric function of its n arguments x'

(b) It is an unbiassed estimate of the parameter θ.

That is, its expectation, over all possible samples of n, must equal θ, or

$$E'(T) = \int T \Phi \, dx' = \theta$$

(the integral being taken over all n variables).

Since $\int \Phi \, dx' = 1$ for all θ, the equation of unbiassedness may be rewritten

$$\int (T - \theta) \Phi \, dx' = 0$$

(c) It has a smaller sampling variance than any other estimate of θ. i.e.

$$V = \int (T - \theta)^2 \Phi \, dx' = \underline{\text{minimum}}$$

<u>3.0.1.</u> <u>Condition of Unbiassedness:</u> Consider first the question of obtaining an unbiassed statistic. All our

information about φ is comprised in the total probability condition

$$\int \varphi \, dx = 1 \quad , \quad \text{or} \quad \int \Phi \, dx' = 1.$$

Since these equations hold for all values of θ, we have

$$\frac{\partial}{\partial \theta} \int \varphi \, dx = 0 \quad ; \quad \frac{\partial}{\partial \theta} \int \Phi \, dx' = 0.$$

or

$$\int \frac{\partial \varphi}{\partial \theta} \, dx = 0 \quad ; \quad \int \frac{\partial \Phi}{\partial \theta} \, dx' = 0 \quad \text{for all } \theta .$$

(Differentiation under the integral sign is legitimate since φ is a continuous function of both x and θ.)

Similarly

$$\int \frac{\partial^r \varphi}{\partial \theta^r} \, dx = 0 \quad \text{for all } \theta ; \int \frac{\partial^r \Phi}{\partial \theta^r} \, dx' = 0 \quad \text{for all } \theta .$$

$$[\, r = 1, 2, 3, \cdots \,] \tag{1}$$

If, now, a function $T(x')$ exists satisfying

$$\int (T - \theta) \, \Phi \, dx' = 0 \tag{2}$$

it follows that $(T - \theta)\Phi$ must equal a linear combination of terms invòlving the $\partial^r \varphi / \partial \theta^r$ and $\partial^r \Phi / \partial \theta^r$ $[r = 1, 2, \cdots]$, with coefficients which may depend on θ but which must not depend on x'. Suppose, then, that there exists a set of numbers $\lambda_1, \lambda_2, \cdots, \mu_2, \mu_3, \cdots$ (involving θ but not x'; zero members of the set are permissible) such that

$$T \equiv \theta + \frac{1}{\Phi}\left[\lambda_1 \frac{\partial \Phi}{\partial \theta} + \lambda_2 \frac{\partial^2 \Phi}{\partial \theta^2} + \lambda_3 \frac{\partial^3 \Phi}{\partial \theta^3} + \cdots \right]$$

$$+ \frac{1}{\Phi}\left[\mu_2 \sum_{i=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)} \cdot \frac{\partial^2 \varphi(x_i|\theta)}{\partial \theta^2} + \mu_3 \sum_{i=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)} \cdot \frac{\partial^3 \varphi(x_i|\theta)}{\partial \theta^3} + \cdots \right] \tag{3}$$

is a function of x' alone. We at once see that

(i) T is symmetrical in its n arguments x'.

(ii) T is an unbiassed estimate of θ in consequence of (i). It may be noted in passing that we do not need to add terms of the form

$$K_{rs} \sum_{i,j=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)\varphi(x_j|\theta)} \cdot \frac{\partial^r \varphi(x_i|\theta)}{\partial\theta^r} \cdot \frac{\partial^s \varphi(x_j|\theta)}{\partial\theta^s}$$

(which retain the property of unbiassedness), for by suitable adjustment of the λ's and μ's these can be included in the scope of (3)

(iii) If a function of x' is not expressible in this form, it is not unbiassed; since any integral of value zero, for all values of θ, can only arise from combinations of expressions such as (1).

We have defined the compound probability density as

$$\Phi(x'|\theta) = \prod_{i=1}^{n} \varphi(x_i|\theta) \qquad (4)$$

This relation is, in fact, a statement of the law of multiplication of independent probabilities. It will now be shown that, unless $\lambda_2 = \lambda_3 = \cdots = 0$ in equation (3), φ will not obey this fundamental law. In other words, we shall show that in the general expression (3) of an unbiassed statistic, we must omit the terms in $\partial^r \Phi/\partial\theta^r$. $[r = 2, 3, \cdots]$.

Suppose that the statement just made is untrue, and that there exist two functions $\lambda_1(\theta)$ and $\lambda_2(\theta)$ such that

$$T \equiv \theta + \frac{1}{\Phi}\left(\lambda_1 \frac{\partial \Phi}{\partial \theta} + \lambda_2 \frac{\partial^2 \Phi}{\partial \theta^2}\right)$$

is dependent on x' alone, Rewrite this as

$$\lambda_2 \frac{\partial^2 \Phi}{\partial \theta^2} + \lambda_1 \frac{\partial \Phi}{\partial \theta} + (\theta - T)\Phi = 0$$

and regard it - T being a given function of x', independent of θ - as an ordinary differential equation for Φ. Solving, as is always possible, by the method of Frobenius, we obtain a solution

$$\Phi = \theta^{c'}\left\{a_0 + a_1'\theta + a_2'\theta^2 + \cdots\right\} \quad (a_0 \neq 0)$$

where c', a_0, a_1' --- are independent of θ, but may involve x'. This will be stressed by writing $c'(x')$. $a_0(x')$, $a_1'(x')$, ----- and the solution becomes

$$\Phi = a_0(x')\,\theta^{c'(x')}\left\{1 + \frac{a_1'(x')}{a_0(x')}\theta + \frac{a_2'(x')}{a_0(x')}\theta^2 + \cdots\right\}$$

Denoting the terms within the bracket by $F(x' \cdot \theta)$, we have

$$\Phi = a_0(x')\,\theta^{c'(x')}\,F(x' \cdot \theta) \qquad (5)$$

Now Φ is a product of n similar factors φ (equation (4)), so that each of its three factors in (5) must be factorisable. That is, there must exist functions $a(x)$, $c(x)$, $f(x \cdot \theta)$ such that

$$a_0(x') = \prod_{i=1}^{n} a(x_i) \quad ; \quad F(x' \cdot \theta) = \prod_{i=1}^{n} f(x_i \cdot \theta)$$

$$c'(x') = \sum_{i=1}^{n} c(x_i)$$

whence we have

$$\varphi(x \mid \theta) = a(x)\,\theta^{c(x)}\,f(x \cdot \theta) \qquad (5')$$

We now use this expression to derive $\partial\Phi/\partial\theta$, $\partial^2\Phi/\partial\theta^2$. Substituting these in the differential equation from which we started, we examine whether our solution is compatible with T ≡ a function of x' alone. Thus

$$\log \varphi = \log a(x) + c(x) \log \theta + \log f(x \cdot \theta)$$

$$\therefore \frac{\partial \varphi}{\partial \theta} = \varphi \left\{ \frac{c(x)}{\theta} + u(x \cdot \theta) \right\}$$ where $u(x \cdot \theta)$ is written for $\frac{1}{f}\frac{\partial f}{\partial \theta}$.

and

$$\frac{\partial^2 \varphi}{\partial \theta^2} = \varphi \left\{ \frac{c(x)}{\theta} + u(x \cdot \theta) \right\}^2 + \varphi \left\{ -\frac{c(x)}{\theta^2} + \frac{\partial u}{\partial \theta} \right\}$$

Also $$\Phi = \exp \left\{ \sum_{i=1}^{n} \log \varphi(x_i \mid \theta) \right\}$$

$$\frac{\partial \Phi}{\partial \theta} = \Phi \sum_{i=1}^{n} \frac{1}{\varphi(x_i \mid \theta)} \frac{\partial \varphi(x_i \mid \theta)}{\partial \theta}$$

$$\frac{\partial^2 \Phi}{\partial \theta^2} = \Phi \left[\left\{ \sum_i \frac{1}{\varphi} \frac{\partial \varphi}{\partial \theta} \right\}^2 + \sum_i \frac{1}{\varphi} \frac{\partial^2 \varphi}{\partial \theta^2} - \sum_i \left(\frac{1}{\varphi} \frac{\partial \varphi}{\partial \theta} \right)^2 \right]$$

Substitute these last three expressions into the equation

$$(T - \theta)\Phi = \lambda_1 \frac{\partial \Phi}{\partial \theta} + \lambda_2 \frac{\partial^2 \Phi}{\partial \theta^2}$$

and replace $\frac{\partial \varphi}{\partial \theta}$, $\frac{\partial^2 \varphi}{\partial \theta^2}$ by their values as noted above.

We obtain, on simplifying and dividing by Φ, which is $\neq 0$,

$$\frac{\lambda_1}{\theta} \sum c(x) + \lambda_1 \sum u(x \cdot \theta) + \frac{\lambda_2}{\theta^2} \sum c(x) \left[\left\{ \sum c(x) \right\} - 1 \right]$$

$$- \lambda_2 \left\{ \sum u(x \cdot \theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} + \frac{2\lambda_2}{\theta} \left\{ \sum c(x) \right\} \left\{ \sum u(x \cdot \theta) \right\}$$

$$\equiv T - \theta \qquad (6)$$

i.e., left hand side $\equiv$ (a function of x') $- \theta$, an identity which must hold for all values of θ and of $x' = (x_1, x_2, \cdots x_n)$

First Case: Suppose $u(x \cdot \theta)$ is genuinely a function of both x and θ.

(i) If $c(x)$ is a function of x and not merely a constant or zero, then in order that (6) be of the required form, we would have, first, to choose $\lambda_1 = \theta$, or $\lambda_2 = \theta$, or both simultaneously, in order to obtain a function of x' alone; and we would be left with one of the following:

$$\lambda_1 \sum u(x\cdot\theta) - \lambda_2 \left\{ \sum u(x\cdot\theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} + \frac{2\lambda_2}{\theta} \sum c(x) \sum u(x\cdot\theta) + \frac{\lambda_2}{\theta^2} \sum c(x) \left[\sum c(x) - 1 \right] \equiv -\theta$$

$$\lambda_1 \sum u(x\cdot\theta) - \lambda_2 \left\{ \sum u(x\cdot\theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} + \frac{2\lambda_2}{\theta} \sum c(x) \sum u(x\cdot\theta) + \frac{\lambda_1}{\theta} \sum c(x) \equiv -\theta$$

$$\lambda_1 \sum u(x\cdot\theta) - \lambda_2 \left\{ \sum u(x\cdot\theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} + \frac{2\lambda_2}{\theta} \sum c(x) \sum u(x\cdot\theta) \equiv -\theta$$

The only term which may equal $-\theta$ is that in $\sum \frac{\partial u}{\partial \theta}$

If so, we also have

$$(\text{function of } x' \text{ and } \theta) \equiv 0$$

and, inspecting the foregoing relations, we see that in each case this can only be satisfied if $\lambda_1 = \lambda_2 = 0$. If $\partial u / \partial \theta$ is not a function of θ alone, none of the above identities is ever satisfied.

(ii) If $c(x)$ is a constant, we might hope to choose λ_1 and/or λ_2 such that the first and/or third terms on the left hand side of (6) equalled $-\theta$. If we so chose, we would require one of the following identities:

$$\lambda_1 \sum u(x \cdot \theta) + \lambda_2 \left\{ \sum u(x \cdot \theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} + \frac{2\lambda_2}{\theta} K \sum u(x \cdot \theta) + \frac{\lambda_1 K}{\theta}$$
$$\equiv \text{function of } x'$$

$$\lambda_1 \sum u(x \cdot \theta) + \lambda_2 \left\{ \sum u(x \cdot \theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} + \frac{2\lambda_2}{\theta} K \sum u(x \cdot \theta) + \frac{\lambda_2}{\theta^2} K(K-1)$$
$$\equiv \text{function of } x'.$$

$$\lambda_1 \sum u(x \cdot \theta) + \lambda_2 \left\{ \sum u(x \cdot \theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} + \frac{2\lambda_2}{\theta} K \sum u(x \cdot \theta) \equiv \text{function of } x'$$

Here K = $\sum c(x)$ = constant.

We observe that none of these identities can subsist (since u is by hypothesis a function of x and θ) however we may restrict $\partial u / \partial \theta$.

(iii) If $c(x)$ is identically zero, (6) becomes

$$\lambda_1 \sum u(x \cdot \theta) + \lambda_2 \left\{ \sum u(x \cdot \theta) \right\}^2 + \lambda_2 \sum \frac{\partial u}{\partial \theta} \equiv \left(\text{function of } x' \right) - \theta \ ,$$

which, again, cannot hold if u is a function of both x, θ.

We conclude that when $u \equiv \frac{1}{f} \frac{\partial f}{\partial \theta} \equiv$ function of both x and θ, there never exist functions $\lambda_1(\theta)$, $\lambda_2(\theta)$ which make $T \equiv \theta + \frac{1}{\Phi}\left(\lambda_1 \frac{\partial \Phi}{\partial \theta} + \lambda_2 \frac{\partial^2 \Phi}{\partial \theta^2} \right)$ dependent on x' alone.

Second Case: Suppose u $(x \cdot \theta)$ degenerates into a function of θ alone, say into $u_1(\theta)$ Identity (6) becomes

$$\frac{\lambda_1}{\theta} \sum c(x) + n\lambda_1 u_1 + \frac{\lambda_2}{\theta^2} \sum c(x) \left[\left\{ \sum c(x) - 1 \right] + \lambda_2 n^2 u_1^2 \right.$$
$$+ \lambda_2 n \frac{du_1}{d\theta} + \frac{2n\lambda_2 u_1}{\theta} \sum c(x) \equiv \left(\text{function of } x' \right) - \theta \qquad (6')$$

Obviously, unless $c(x)$ is genuinely a function of x, and not a mere constant or zero, this identity cannot be satisfied. Again, even when $c(x)$ is a function of x, we also require

$$\frac{2n\lambda_2 u_1}{\theta} \sum c(x) = 0$$

The solution $u_1 = 0$ is unacceptable, on referring back to (6). Therefore $\lambda_2 = 0$, which we see by inspection is an acceptable solution of $(6')$.

Third Case: Suppose that $u(x \cdot \theta)$ degenerates into a function of x alone. Then, by inspection, we can never satisfy (6).

The foregoing enumerates completely the different possible ways in which the identity (6) might be satisfied. We have thus established that there is no probability distribution φ for which there exist non-zero functions $\lambda_1(\theta)$, $\lambda_2(\theta)$ which make

$$T \equiv \theta + \frac{1}{\Phi}\left(\lambda_1 \frac{\partial \Phi}{\partial \theta} + \lambda_2 \frac{\partial^2 \Phi}{\partial \theta^2}\right)$$

dependent on x' alone.

We next investigate whether, for a distribution φ satisfying the multiplication law of probabilities, there could exist three or more functions $\lambda_1(\theta), \lambda_2(\theta), \lambda_3(\theta), \cdots$ which would make

$$T \equiv \theta + \frac{1}{\Phi}\left\{\lambda_1 \frac{\partial \Phi}{\partial \theta} + \lambda_2 \frac{\partial^2 \Phi}{\partial \theta^2} + \lambda_3 \frac{\partial^3 \Phi}{\partial \theta^3} + \cdots\right\}$$

dependent on x' alone. By repeating the foregoing analysis step by step, it is found no such distribution φ exists unless

$$\lambda_2 = \lambda_3 = \cdots\cdots = 0 .$$

We have thus demonstrated the following - while the general unbiassed statistic is of the form (3), the multiplication law of independent probabilities forces us to reject all cases where $\lambda_2, \lambda_3, \cdots$ are not zero. We are therefore left with the cases where some or all of $\lambda_1, \mu_2, \mu_3, \cdots$ exist, such that

$$T \equiv \theta + \frac{1}{\Phi}\lambda_1 \frac{\partial \Phi}{\partial \theta} + \frac{1}{\Phi}\left[\mu_2 \sum_{i=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^2 \varphi(x_i|\theta)}{\partial \theta^2} + \mu_3 \sum_{i=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^3 \varphi(x_i|\theta)}{\partial \theta^3} + \cdots\cdots\right]$$

is a function of x' alone. It is clear that such a function (however many of the $\lambda_1, \mu_2, \mu_3, \cdots$ exist) <u>is</u> compatible with equation (4) — for, in fact, it is built up from the original φ and its derivatives.

<u>3.0.2</u> <u>Minimum Variance</u>:- We turn to consider the condition that

$$V = \int (T-\theta)^2 \Phi \, dx'$$

be a minimum. Suppose that $T(x')$ is altered by a small amount to $T(x') + \epsilon f(x')$ where $f(x')$ is an arbitrary function of x', and ϵ is a small arbitrary constant. The corresponding variance is

$$V + \delta V = \int (T + \epsilon f - \theta)^2 \Phi \, dx' \quad .$$

Subtracting, and taking ϵ so small that terms in ϵ^2 are negligible, we have

$$\delta V = 2\epsilon \int f(x')(T-\theta)\Phi \, dx' .$$

In order that T may be a statistic which makes the variance stationary, we require $\delta V = 0$, <u>or</u>

$$\int f(x')(T-\theta)\Phi \, dx' = 0 \quad .$$

It is apparent that, if Φ is a (compound) probability distribution, there is no function T which satisfies the last equation for <u>every possible</u> arbitrary function $f(x')$. We must therefore, to progress at all, define some class of arbitrary functions f and, with this restriction, try to satisfy $\delta V = 0$.

The class of arbitrary functions which we shall consider is

$$f(x') = \sum_r \frac{a_r}{\Phi} \frac{\partial^r \Phi}{\partial \theta^r} + \sum_s b_s u_s \tag{7}$$

where

(i) $\frac{1}{\Phi} \frac{\partial^r \Phi}{\partial \theta^r}$ is evaluated for the actual population value of the parameter θ. [In <u>practice</u>, we may not know this particular value; but for a theoretical investigation this is immaterial. Such a value exists by hypothesis, and we consider those functions of x' which can be built up from $\frac{1}{\Phi} \frac{\partial^r \Phi}{\partial \theta^r}$ $[r = 1, 2, \cdots]$ when θ assumes this value.]

(ii) u_s is any solution of the integral equation

$$\int u(x') \, \Phi(x'|\theta) \, dx' = 1.$$

We shall prove shortly that this equation has an infinite number of solutions.

(iii) the terms a_r, b_s are arbitrary numerical constants.

(iv) the summations $\sum_r$, $\sum_s$ cover as many terms as one pleases.

Since our statistic T is to be unbiassed, it can be expressed in the form

$$(T-\theta)\Phi = \lambda_1 \frac{\partial \Phi}{\partial \theta} + \mu_2 \sum_{i=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^2 \varphi(x_i|\theta)}{\partial \theta^2} + \mu_3 \sum_{i=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^3 \varphi(x_i|\theta)}{\partial \theta^3} + \cdots\cdots$$

Substituting in δV, we obtain

$$\delta V = 2\epsilon \int f(x') \left[\lambda_1 \frac{\partial \Phi}{\partial \theta} + \mu_2 \sum_{i=1}^{n} \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^2 \varphi(x_i|\theta)}{\partial \theta^2} + \cdots \right] dx'$$

$$= 2\epsilon \lambda_1 \delta V_1 + 2\epsilon \mu_2 \delta V_2 + \cdots\cdots \qquad (8)$$

where

$$\delta V_1 = \int f(x') \frac{\partial \Phi}{\partial \theta} dx' \; ; \; \delta V_2 = \sum_{i=1}^{n} \int \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^2 \varphi(x_i|\theta)}{\partial \theta^2} f(x') dx' \; ; \; \text{etc.}$$

When $f(x')$ is of the form (7)

$$\delta V_1 = \frac{\partial}{\partial \theta_1} \int f(x') \Phi \, dx'$$

$$= \frac{\partial}{\partial \theta} \int \sum_r a_r \frac{\partial^r \Phi}{\partial \theta^r} dx' + \frac{\partial}{\partial \theta} \int \sum_s b_s u_s \Phi \, dx'$$

$$= \frac{\partial}{\partial \theta} \sum_r a_r \int \frac{\partial^r \Phi}{\partial \theta^r} dx' + \frac{\partial}{\partial \theta} \sum_s b_s \int u_s \Phi \, dx'$$

$$= \sum_s b_s \frac{\partial}{\partial \theta} \int u_s \Phi \, dx' \; , \; \text{since} \int \frac{\partial^r \Phi}{\partial \theta^r} dx' = 0$$

$$= 0 \qquad \text{since} \int u_s \Phi \, dx' = 1 .$$

Again

$$\delta V_2 = \sum_{i=1}^{n} \int \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^2 \varphi(x_i|\theta)}{\partial \theta^2} \sum_r \frac{a_r}{\Phi} \frac{\partial^r \Phi}{\partial \theta^r} dx' + \sum_{i=1}^{n} \int \frac{\Phi}{\varphi(x_i|\theta)} \frac{\partial^2 \varphi(x_i|\theta)}{\partial \theta^2} \sum_s b_s u_s \, dx'$$

Clearly this does not in general reduce to zero.

Similarly $\delta V_3 \neq 0$, $\delta V_4 \neq 0$, $\cdots\cdots$

Hence referring to (8), the necessary and sufficient condition for $\delta V = 0$ is

$$\mu_2 = \mu_3 = \text{------} = 0 .$$

Finally, then, for variations $f(x')$ of the

p70 type (7), an unbiassed statistic exists, which makes the variance stationary, if a function $\lambda_1(\theta)$ can be found such that

$$T \equiv \theta + \frac{\lambda_1}{\Phi}\frac{\partial \Phi}{\partial \theta} \qquad (9)$$

is a function of x' alone. The statistic in question is T itself.

<u>Nature of The Stationary Value of the Variance</u> :-

When $T(x')$ is changed by an arbitrary amount $\epsilon f(x')$, the variance V is changed to

$$\int (T + \epsilon f - \theta)^2 \Phi \, dx' \; ;$$

when T is given by (9), the first variation is zero, and the second variation is

$$\delta_2 V = \epsilon^2 \int f^2 \Phi \, dx' \, .$$

Since $\Phi \geqslant 0$ everywhere (being a probability distribution),

$$\delta_2 V > 0 .$$

Our value of T therefore makes the variance a <u>minimum</u>.

3.0.4 <u>Comments</u> :- Our problem at first sight may have seemed one which might be tackled by the calculus of variations in a straightforward way. We have had, however, to proceed by a route very different from that normally followed in this branch of analysis. For instance, we had to minimise an integral (the variance) subject to a condition (unbiassedness); we might therefore have expected the concept of the Lagrangian multiplier to be of service. While a term $\lambda_1(\theta)$ enters into our final solution, its mode of introduction suggests that it would be somewhat artificial to describe it as a Lagrangian coefficient.

A more important point concerns the restriction imposed on our arbitrary function $f(x')$. In the normal procedure of the variational calculus, one finds that the first variation δV can be expressed in the form $\int f(x') G(x')\,dx'$ where f is a completely arbitrary function. This integral is always zero if $G(x') = 0$ everywhere; and $G = 0$ generally represents a differential equation from which the unknown function is determined. In our problem, however, G is replaced by $T(x') - \theta$, and the equating of this to zero does not lead to any solution; it would, in fact, give a quite meaningless relation, since we wish to find $T(x')$ as a function of the observations $x' \left(= x_1, x_2 \cdots, x_n\right)$. In other words, in the ordinary calculus of variations one determines a curve along which the integrand of the expression for δV is everywhere zero. This was not, however, possible in our problem, and we had, accordingly, first to restrict the permissible forms of the arbitrary $f(x')$ and then to make zero the integral which expresses δV, not the integrand of this integral.

From a practical point of view, it might be feared that the restriction on $f(x')$ might limit the scope of the present method of estimation. Such a fear is, however, groundless. It will be shown in later Sections (3·6 and 3.8) that an unbiassed statistic of minimum variance exists for a wide and very important class of distributions, and that this statistic is identical with that yielded by the method of Maximum Likelihood. For this class of distributions, we find that the first part of the expression for $f(x')$ viz.,

$$\sum_r \frac{a_r}{\Phi} \cdot \frac{\partial^r \Phi}{\partial \theta^r}$$

(evaluated for the population value of θ) represents any function of the unbiassed statistic of minimum variance which is capable of expansion in a Maclaurin series.

3.1 Non-Uniqueness of The Unbiassed Statistic of Minimum Variance:- Consider a distribution $\varphi(x/\theta)$ for which an unbiassed statistic T of minimum variance exists. Then there is a function $\lambda_1(\theta)$, not involving the observations x', such that

$$T = \theta + \frac{\lambda_1(\theta)}{\Phi} \frac{\partial \Phi}{\partial \theta}$$

Let us enquire whether any other statistic could have the same properties. If a second such statistic S does exist, we have the following relations:

$$\int T \Phi \, dx' = \theta \quad ; \quad \int S \cdot \Phi \, dx' = \theta$$

$$\int (T-\theta)^2 \Phi \, dx' = V_{min}(\theta) \quad ; \quad \int (S-\theta)^2 \Phi \, dx' = V_{min}(\theta).$$

Hence $\int T^2 \Phi \, dx' = V_{min}(\theta) + \theta^2$

$\int S^2 \Phi \, dx' = V_{min}(\theta) + \theta^2$

and $\int (T^2 - S^2) \Phi \, dx' = 0$ (for all values of θ)

Writing $g(x') \equiv T^2 - S^2$, our question can be put - has the integral equation $\int g(x') \Phi(x'|\theta) \, dx' = 0$ a non-zero continuous solution or solutions g? The answer is that this equation does have an infinite number of continuous solutions.

Let $\Phi^{\times}(x'|\theta')$ denote a continuous, finite function of $x' = (x_1, x_2, \cdots x_n)$ and $\theta' = (\theta_1, \theta_2, \cdots \theta_n)$ such that, when the n parameters θ' are all equal to a value θ, $\Phi^{\times}$ assumes the value $\Phi(x'|\theta)$. We denote this condition by

$$\left[\Phi^{\times}(x'|\theta')\right]_{\theta_1 = \theta_2 = \cdots = \theta_n = \theta} = \Phi(x'|\theta).$$

Since Φ is a product of n factors $\varphi(x_i|\theta)$, it is reasonable to stipulate that $\Phi^{\times}$ should be a similar product. We may therefore, with little real loss of generality, take

$$\Phi^{\times}(x'|\theta') = \Phi(x'|\theta') = \prod_{i=1}^{n} \varphi(x_i|\theta_i)$$

This satisfies the above mentioned condition.

Let $v^{\times}(\theta')$ be a continuous function of $\theta' = (\theta_1, \theta_2, \cdots \theta_n)$ such that

$$\left[v^{\times}(\theta')\right]_{\theta_1 = \theta_2 = \cdots = \theta_n = \theta} = V_{min}(\theta) + \theta^2.$$

Even for our present statistical application, this condition does not uniquely determine our function v^{*}. It might, for instance, be taken as

$$v^{*}(\theta') = \frac{1}{n}\left\{\sum_i V_{min}(\theta_i)\right\} + \frac{1}{n}\sum_i \theta_i^{2}$$

or

$$v^{*}(\theta') = \frac{1}{n}\sum_i \frac{V_{min}(\theta_i)\, w_i(\theta_i)}{w_1(\theta_1) + w_2(\theta_2) \cdots + w_n(\theta_n)} + \frac{1}{n}\sum_i \theta_i^{2}$$

and so on.

Choose any two of the possible functions $v^{*}(\theta')$, say $v_1^{*}(\theta')$, $v_2^{*}(\theta')$ and consider the homogeneous linear integral equations

$$\int_a^b g_1(x')\,\Phi(x'/\theta')\,dx' = v_1^{*}(\theta')$$
$$\int_a^b g_2(x')\,\Phi(x'/\theta')\,dx' = v_2^{*}(\theta').$$

where a, b are the limits of the variates $(x_1, x_2, \cdots x_n)$. We know that such equations have each a unique solution of class L^{2}, and we know how to obtain it. If the range (a, b) is infinite, we transform the variables so that it becomes finite. The kernel being a probability distribution, the transformed kernel has no singularities). We "symmetrise" the kernel, and, as explained in Chapter Two, proceed to determine the complete set of characteristic numbers and of characteristic functions of the symmetrised kernel. Let these be $\{\lambda\}$ and $\{u(x')\}$ respectively. Then we find the respective solutions of class L^{2} of the two equations are

$$g_1(x') = \sum_j \lambda_j c_{1j} u_j(x')$$

$$g_2(x') = \sum_j \lambda_j c_{2j} u_j(x')$$

where

$$\left.\begin{aligned} c_{pj} &= \int_a^b w_p(\xi') u_j(\xi') d\xi' \\ w_p(\xi') &= \int_a^b v_p^x(\theta') \Phi(\xi'/\theta') d\theta' \end{aligned}\right\} \quad p = 1, 2.$$

Assume that the series $\sum_j \lambda_j^2 c_{pj}^2$ $(p = 1, 2)$ are convergent, so that the solutions written above are valid. Since these solutions satisfy the original integral equations for all values of $\theta' = (\theta_1, \theta_2, \cdots \theta_n)$ they satisfy them in particular when $\theta_1 = \theta_2 = \cdots = \theta_n = \theta$ (say). That is

$$\int_a^b g_1(x') \left[\Phi(x'/\theta')\right]_{\theta_1 = \cdots = \theta_n = \theta} \cdot dx' = \left[v_1^x(\theta')\right]_{\theta_1 = \cdots = \theta_n = \theta}$$

$$\int_a^b g_2(x') \left[\Phi(x'/\theta')\right]_{\theta_1 = \cdots = \theta_n = \theta} \cdot dx' = \left[v_2^x(\theta')\right]_{\theta_1 = \cdots = \theta_n = \theta}$$

or

$$\int_a^b g_1(x') \Phi(x'/\theta) dx' = V_{min}(\theta) + \theta^2$$

$$\int_a^b g_2(x') \Phi(x'/\theta) dx' = V_{min}(\theta) + \theta^2$$

Subtracting,

$$\int_a^b \{g_1(x') - g_2(x')\} \Phi(x'/\theta) dx' = 0,$$

and this holds for any value of θ. As $g_1 \not\equiv g_2$, we have obtained a non-zero solution, of class L^2, of the equation $\int_a^b g(x') \Phi(x'/\theta) dx' = 0$. Corresponding to the infinite number of functions $v^x(\theta')$ which can be chosen, there is an infinite number of solutions, of class L^2, of this equation. By a suitable selection of the $v^x(\theta')$ we can, moreover, obtain a sub-set of solutions which are not only of class L^2, but are also continuous. Thus there is an infinite number of statistics S which have the same second moment about the origin as has T, the unbiassed statistic of minimum variance [i.e.

$$\int S^2 \Phi\, dx' = \int T^2 \Phi\, dx' = V_{min}(\theta) + \theta^2 \,]$$

We now examine whether any of these statistics S also have the same first moment about the origin as has T; i.e. whether

$$\int S \Phi\, dx' = \theta$$

where S is any function satisfying $\int S^2 \Phi\, dx' = V_{min}(\theta) + \theta^2$.

This question is most readily answered by consideration of two sets of algebraic equations whose limits are the integral equations we are studying. Let the range (a, b) of the variate x_i be sub-divided into $t-1$ portions by the values

$$a = x_{i_1},\ x_{i_2},\ x_{i_3},\ \cdots\cdots\ x_{i_t} = b \qquad (i = 1, 2, \cdots n)$$

Subdivide the range of variation of θ into $t-1$ portions by the values $\theta_{(1)},\ \theta_{(2)},\ \cdots\ \ \theta_{(t)}$

Consider now the t equations

$$\sum_{i,j,p\cdots=1}^{t} S\left(x_{1_i}, x_{2_j}, \cdots x_{n_p}\right) \Phi\left(x_{1_i}, x_{2_j}, \cdots x_{n_p} \mid \theta_{(\omega)}\right) h^n = \theta_{(\omega)}$$

and the t equations

$$\sum_{i,j,\cdots p=1}^{t} S^2\left(x_{1_i}, x_{2_j}, \cdots x_{n_p}\right) \Phi\left(x_{1_i}, x_{2_j}, \cdots x_{n_p} \mid \theta_{(\omega)}\right) h^n = V_{min}(\theta_{(\omega)}) + \theta_{(\omega)}^2$$

$$[\omega = 1, 2, \cdots\ t\ ;\ h = x_{i_2} - x_{i_1} = x_{i_3} - x_{i_2} = \cdots\ ]$$

These two sets represent $2t$ algebraic equations in t^n unknown quantities $S\left(x_{1_i}, x_{2_j}, \cdots x_{n_p}\right) \cdot [i = 1, 2, \cdots t;\ \cdots\ p = 1, 2, \cdots t]$. In general, therefore, there is an infinite number of values of each unknown which satisfies the foregoing equations. As t tends to infinity, the two sets of equations tend respectively to

$$\int S \Phi \, dx' = \theta$$
$$\int S^2 \Phi \, dx' = V_{min}(\theta) + \theta^2 .$$

There is therefore an infinite number of functions that satisfy this pair of equations. Hence, in addition to $T = \theta + \frac{\lambda_1(\theta)}{\Phi} \frac{\partial \Phi}{\partial \theta}$, there are other statistics with the same variance and mean.

3.1.1. We have relied, just now, on the fact that a result, which holds for a set of algebraic equations, remains valid as the number of equations increases indefinitely (the set tending to a single integral equation). This is of course true, as is fully proved in the Theory of Functionals.

3.1.2 In discussing in Section 3.0.2 the nature of the arbitrary variation, it was asserted that the equation

$$\int u(x') \, \Phi(x'/\theta) \, dx' = 1$$

possessed an infinite number of solutions. This we have now proved. To obtain one of the solutions, $\Phi(x'/\theta)$ is replaced by $\Phi(x'/\theta')$, and any continuous function $v^{\times}(\theta')$ of n variables is chosen, with the property

$$\left[v^{\times}(\theta') \right]_{\theta_1 = \theta_2 = \cdots = \theta_n = \theta} = 1 .$$

Corresponding to each such $v^{\times}(\theta')$ there is a linear integral equation which has one solution of class L^2.

3.2 A Generalisation of The Problem of Estimation:- The non-uniqueness of the unbiassed statistic of minimum variance might be awkward, were it not that a simple generalisation of the initial problem may be made, which does lead to a unique solution.

Thus far, our population has been specified by a coefficient θ, of unknown but fixed value. Suppose, however, that θ is liable to vary, in some unknown manner, between one observation and the next. The probability that the i^{th} observation lies between $x_i \pm \frac{1}{2} dx_i$ is $\varphi(x_i | \theta_i)\, dx_i$ where θ_i may, for all we know, differ from θ_{i-1} and θ_{i+1}.

Consider the problem of estimating the mean of the parameters

$$\bar{\theta} = \frac{1}{n}\left\{ \theta_1 + \theta_2 + \cdots + \theta_n \right\},$$

from a random sample of n observations $x' = (x_1, x_2, \cdots x_n)$. The criteria to be applied are, as before, those of unbiassedness and minimum variance. We therefore seek to determine a function of the observations $T(x')$, which is symmetric in its n arguments, and is such that

$$\int (T - \bar{\theta})\, \Phi(x'|\theta')\, dx' = 0$$

$$V = \int (T - \bar{\theta})^2\, \Phi(x'|\theta')\, dx' = \text{minimum}$$

Here $\Phi(x'|\theta') = \prod_{i=1}^{n} \varphi(x_i | \theta_i)$,

and we have, in addition, the total probability condition.

$$\int \Phi(x'|\theta')\, dx' = 1 \quad \text{for all} \quad \theta' = (\theta_1, \theta_2 \cdots \theta_n)$$

The analysis of Section 3.0 may be applied with only minor modifications. Thus the condition of unbiassedness is satisfied, subject to the total probability condition, if there exist functions $\lambda_1(\theta_i), \lambda_2(\theta_i), \cdots ; \mu_2(\theta_i) \cdots$ which make

$$T \equiv \bar{\theta} + \frac{1}{\Phi(x'|\theta')}\left[\sum_{i=1}^{n} \lambda_1(\theta_i)\frac{\partial \Phi(x'|\theta')}{\partial \theta_i} + \sum_{i=1}^{n} \lambda_2(\theta_i)\frac{\partial^2 \Phi(x'|\theta')}{\partial \theta_i^2} + \cdots\cdots + \sum_{i=1}^{n} \mu_2(\theta_i)\frac{\Phi(x'|\theta')}{\varphi(x_i|\theta_i)} \frac{\partial^2 \varphi(x_i|\theta_i)}{\partial \theta_i^2} + \cdots\right]$$

dependent on x' alone. (The λ's and μ's are functions of θ alone, not of x'.)

In order that φ should obey the multiplicative law,

$$\Phi(x'|\theta') = \prod_{i=1}^{n} \varphi(x_i|\theta_i)$$

it is necessary that $\lambda_2 = \lambda_3 = \cdots = 0$.

By altering T by a small arbitrary amount $\epsilon f(x')$ we derive the first variation of the value of the variance as

$$\delta V = 2\epsilon \int f(x')\left(T(x') - \bar{\theta}\right) \Phi(x'|\theta')\, dx'$$

We restrict ourselves to variations $f(x')$ which are of the form

$$f(x') = \sum_r \left\{ \sum_{i=1}^{n} \frac{a_r}{\Phi(x'|\theta')} \frac{\partial^r \Phi(x'|\theta')}{\partial \theta_i^r} \right\} + b u$$

Here, (i) the terms in the sum are to be evaluated for the actual population values of the $\theta_1, \theta_2, \cdots \theta_n$ assumed in the course of the observations.

(ii) u is ~~any~~ the solution of $\int u(x')\Phi(x'|\theta')\, dx' = 1$.

(iii) the a's and b are arbitrary numerical constants, and the summations $\sum_r$ covers as many terms as one pleases.

As in Section 3.0.2, we find that, with $f(x')$ thus restricted, the necessary and sufficient condition for $\delta V = 0$ is $\mu_2 = \mu_3 = \cdots \cdots = 0$.

Hence, provided that we can find a function (of θ only) $\lambda_i(\theta)$ such that

$$T \equiv \bar{\theta} + \sum_{i=1}^{n} \lambda_i(\theta_i)\left(\frac{1}{\Phi(x'|\theta')}\frac{\partial \Phi(x'|\theta')}{\partial \theta_i}\right) \tag{10}$$

is dependent on x' alone, an unbiassed estimate of $\bar{\theta}$ exists for which the variance is stationary. It is readily seen that the stationary value is a minimum, and that T is the estimate in question.

3.2.1 The distribution just considered, in which the parameter is liable to vary from one observation to the next, bears to the usual distribution (specified by a single fixed value of the coefficient θ), the same relation as does Poisson's binomial distribution to Bernouillis' binomial distribution. We shall find it a convenience to attach the adjectives Poissonian and Bernouillian to the sample, rather than to the distribution. Let a population be specified by the probability distribution $\varphi(x|\theta)$. If, in the course of making n observations at random, the value of θ is known not to vary, we shall say that we have a Bernouillian sample. If, during the observations, the value of θ is liable to alter, we describe the sample as a Poissonian sample.

3.2.2 Uniqueness of The Unbiassed Statistic of Minimum Variance in a Poissonian Sample:- Suppose that, in a Poissonian sample, there exists another unbiassed statistic S, estimating $\bar{\theta}$, which has the same variance as

$$T = \bar{\theta} + \sum_{i=1}^{n} \lambda_1(\theta_i) \cdot \frac{1}{\Phi(x'|\theta')} \cdot \frac{\partial \Phi(x'|\theta')}{\partial \theta_i} ;$$

we then have

$$\int T \Phi(x'|\theta')\, dx' = \bar{\theta} \quad ; \quad \int S \Phi(x'|\theta')\, dx' = \bar{\theta}$$

$$\int (T-\bar{\theta})^2 \Phi(x'|\theta')\, dx' = V_{min}(\theta') ; \int (S-\bar{\theta})^2 \Phi(x'|\theta')\, dx' = V_{min}(\theta') .$$

It follows that

$$\int T^2 \Phi(x'|\theta')\, dx' = V_{min}(\theta') + \bar{\theta}^2$$

$$\int S^2 \Phi(x'|\theta')\, dx' = V_{min}(\theta') + \bar{\theta}^2$$

whence $\int (T^2 - S^2)\Phi(x'|\theta')\, dx' = 0$ for all values of $\theta' = (\theta_1, \theta_2, \cdots \theta_n)$

Now, from Chapter Two, we know that the only continuous solution of this integral equation is

$$T^2 - S^2 \equiv 0 , \quad \text{or} \quad T \equiv S .$$

Hence T is the only continuous function which estimates $\bar{\theta}$ without bias and which has minimum sampling variance.

We may express this result by saying that an unbiassed statistic of minimum variance is unique for Poissonian samples, but not for Bernouillian samples.

3.2.3 This property of uniqueness justifies a certain step in the foregoing analysis. Using the total probability condition, it was stated that the general form of an unbiassed statistic was

$$T \equiv \bar{\theta} + \sum_{i=1}^{n} \frac{\lambda_1(\theta_i)}{\Phi(x'|\theta')} \frac{\partial \Phi(x'|\theta')}{\partial \theta_i} + \sum_{i=1}^{n} \frac{\mu_2(\theta_i)}{\varphi(x_i|\theta_i)} \frac{\partial^2 \varphi(x_i|\theta_i)}{\partial \theta_i^2} + \cdots\cdots;$$

While this is certainly sufficient, we refrained from any attempt to show that it was the most general form. Such an attempt is unnecessary, however, since the statistic we have obtained (in equation (10)) is the only one possessing the required properties in a Poissonian sample.

3.3 Variance of The Unbiassed Statistic of Minimum Variance:-

3.3.1 In Poissonian Samples:- The actual variance of

our statistic

$$T = \bar{\theta} + \sum_{i=1}^{n} \frac{\lambda_1(\theta_i)}{\Phi(x'|\theta')} \cdot \frac{\partial \Phi(x'|\theta')}{\partial \theta_i}$$ is

$$V(\theta') = \int (T - \bar{\theta})^2 \Phi \, dx' = \int \left\{ \sum_{i=1}^{n} \lambda_1(\theta_i) \frac{\partial \Phi(x'|\theta')}{\partial \theta_i} \right\}^2 \frac{dx'}{\Phi(x'|\theta')}$$

$$= \sum_{i=1}^{n} \lambda_1^2(\theta_i) \int \left(\frac{\partial \Phi(x'|\theta')}{\partial \theta_i} \right)^2 \frac{dx'}{\Phi(x'|\theta')} + \sum_{i,j=1}^{n}{}' \lambda_1(\theta_i)\lambda_1(\theta_j) \int \frac{\partial \Phi}{\partial \theta_i} \frac{\partial \Phi}{\partial \theta_j} \frac{dx'}{\Phi}$$

where, in $\sum'$, the terms $i = j$ are omitted.

Now

$$\frac{1}{\Phi} \frac{\partial \Phi}{\partial \theta_i} \frac{\partial \Phi}{\partial \theta_j} = \frac{\partial^2 \Phi}{\partial \theta_i \partial \theta_j} - \Phi \frac{\partial^2 \log \Phi}{\partial \theta_i \partial \theta_j} \quad \begin{pmatrix} i = 1, 2, \cdots n \\ j = 1, 2 \cdots n \end{pmatrix}$$

and

$$\int \frac{\partial^2 \Phi}{\partial \theta_i \partial \theta_j} dx' = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \int \Phi(x'|\theta') dx' = 0 \text{ for all } i, j.$$

Also

$$\log \Phi(x'|\theta') = \sum_{i=1}^{n} \log \varphi(x_i|\theta_i).$$

so

$$\frac{\partial^2 \log \Phi(x'|\theta')}{\partial \theta_i^2} = \frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^2} \quad (i = 1, 2, \cdots n)$$

$$\frac{\partial^2 \log \Phi(x'|\theta')}{\partial\theta_i \, \partial\theta_j} = 0 \qquad (i \neq j)$$

Substituting these results into the expression for $V(\theta')$ we obtain

$$V(\theta') = -\sum_{i=1}^{n} \lambda_1^2(\theta_i) \int \Phi(x'|\theta') \frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^2} \, dx'$$

On writing this n-fold integral in full, we see that the coefficient of $\lambda_1^2(\theta_i)$ is

$$\int \cdots \int \varphi(x_1|\theta_1) \cdots\cdots \varphi(x_n|\theta_n) \frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^2} \, dx_1 \ldots dx_n$$

$$= \left\{ \int \varphi(x_i|\theta_i) \frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^2} \, dx_i \right\} \left\{ \int \varphi(x|\theta) \, dx \right\}^{n-1}$$

$$= \int \varphi(x_i|\theta_i) \frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^2} \, dx_i$$

Hence $V(\theta') = -\sum_{i=1}^{n} \lambda_1^2(\theta_i) \int \varphi(x_i|\theta_i) \frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^2} \, dx_i$ (11)

We shall obtain another expression for the variance. Differentiate

$$T - \bar{\theta} = \sum_{i=1}^{n} \lambda_1(\theta_i) \frac{\partial \log \Phi(x'|\theta')}{\partial \theta_i} = \sum_{i=1}^{n} \lambda_1(\theta_i) \frac{\partial \log \varphi(x_i|\theta_i)}{\partial \theta_i}$$

with respect to θ_i. Remembering that $\bar{\theta} = \frac{1}{n} \sum_i \theta_i$, we have

$$-\frac{1}{n} = \lambda_1(\theta_i) \frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^2} + \frac{\partial \log \varphi(x_i|\theta_i)}{\partial \theta_i} \frac{d\lambda_1(\theta_i)}{d\theta_i} \qquad [i = 1, 2, \ldots n].$$

Multiply by $\varphi(x_i|\theta_i)$ and integrate with respect to x_i. We obtain, on simplifying by means of the total probability condition

$$-\frac{1}{n} = \lambda_1(\theta_i)\int \varphi(x_i|\theta_i)\,\frac{\partial^2 \log \varphi(x_i|\theta_i)}{\partial \theta_i^{\,2}}\,dx_i\,.$$

Substitution in (11) gives at once

$$V(\theta') = \frac{1}{n}\sum_{i=1}^{n} \lambda_1(\theta_i)\,. \tag{12}$$

<u>3.3.2</u> <u>In Bernouillian Samples</u>:- All the formulae developed for Poissonian samples in Section 3.2 reduce, when $\theta_1 = \theta_2 = \cdots = \theta_n = \theta$, to the corresponding formulae for Bernouillian samples. This must hold good, too, for the results (11) and (12), so that we have

$$V_B(\theta) = -\lambda_1^2(\theta)\int \Phi(x'|\theta)\,\frac{\partial^2 \log \Phi(x'|\theta)}{\partial \theta^2}\,dx' \tag{13}$$

or $$V_B(\theta) = \lambda_1(\theta)\,, \tag{14}$$

where V_B denotes the variance, in a Bernouillian sample, of the unbiassed statistic $T = \theta + \frac{\lambda_1(\theta)}{\Phi}\frac{\partial \Phi}{\partial \theta}$ of minimum variance.

The expressions (13) and (14) can of course be obtained directly, by following the analysis of Section 3.3.1 step by step.

<u>3.3.3</u> <u>Comparison of Variance in Poissonian and Bernouilliam Samples</u>:- Let us compare the variance of our statistic T in a Poissonian sample of n, with the variance of the corresponding statistic in a Bernouillian sample from a population specified by the coefficient $\bar{\theta}$.

If the function λ_1 possesses first and second derivatives, we have

$$\lambda_1(\theta_i) = \lambda_1\{\bar{\theta} + (\theta_i - \bar{\theta})\}$$

$$= \lambda_1(\bar{\theta}) + (\theta_i - \bar{\theta})\frac{d\lambda_1}{d\bar{\theta}} + \frac{(\theta_i - \bar{\theta})^2}{2}\frac{d^2\lambda_1\{\bar{\theta} + \rho_i(\theta_i - \bar{\theta})\}}{d\theta_i^2}$$

where $0 \le \rho_i < 1$ and where $\dfrac{d\lambda_1}{d\bar{\theta}} = \left[\dfrac{d\lambda_1(\theta)}{d\theta}\right]_{\theta=\bar{\theta}}$

Hence, from (12)

$$V(\theta') = \lambda_1(\bar{\theta}) + \frac{1}{n}\frac{d\lambda_1}{d\bar{\theta}}\sum_{i=1}^{n}(\theta_i - \bar{\theta}) + \frac{1}{2n}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2\frac{d^2\lambda_1\{\bar{\theta} + \rho_i(\theta_i - \bar{\theta})\}}{d\theta_i^2}$$

Since the variance in the Bernouillian sample is $V_B(\bar{\theta}) = \lambda_1(\bar{\theta})$ and since $\sum_i(\theta_i - \bar{\theta}) = 0$, we have

$$V(\theta') - V_B(\bar{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2\frac{d^2\lambda_1\{\bar{\theta} + \rho_i(\theta_i - \bar{\theta})\}}{d\theta_i^2}.$$

Three special cases may be noted

(i) $\dfrac{d^2\lambda_1}{d\theta^2} > 0$ for all θ; the variance is greater in a Poissonian sample than in a Bernouillian sample from a population specified by parameter $\bar{\theta}$.

(ii) $\dfrac{d^2\lambda_1}{d\theta^2} < 0$ for all θ; the variance is less in a Poissonian sample.

(iii) $\dfrac{d^2\lambda_1}{d\theta^2} = 0$ everywhere; the variance is the same in the two kinds of sample.

Example: Choose $\lambda_1(\theta) = n\theta(1-\theta)$. (This is the well known expression for the variance in the binomial distribution. It can be shown that, though the preceding analysis deals with continuous distributions, the results remain valid for discrete probability distributions)

We have $d^2\lambda_1/d\theta^2 = -2n$. The variance is therefore less in a Poissonian sample than in a Bernouillian sample, (whose parameter is the mean of the parameters in the Poissonian case). Moreover the general expression for $V(\theta') - V_B(\bar{\theta})$ shows that the variance is less by an amount

$$\frac{1}{2n}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2(2n) = \sum_{i=1}^{n}\theta_i^2 - n\bar{\theta}^2 .$$

These results are well known in statistical theory (See, for example, Aitken, "Statistical Mathematics" p.52).

3.4 Distributions which Admit An Unbiassed Statistic of Minimum Variance:- An unbiassed statistic T, of minimum variance, in a Bernouillian sample, has the form

$$T = \theta + \frac{\lambda_1(\theta)}{\Phi}\frac{\partial\Phi}{\partial\theta}$$

or

$$\frac{\partial \log \Phi}{\partial\theta} = \frac{T-\theta}{\lambda_1(\theta)}$$

We may regard this (T being given) as a differential equation in Φ. By solving it, we shall find the form of distributions Φ (and hence of φ) which admit an unbiassed statistic of minimum variance.

Write

$$\int_0^\theta \frac{d\theta}{\lambda_1(\theta)} = F_1(\theta) \quad ; \int_0^\theta F_1(\theta)\,d\theta = \mathcal{G}_1(\theta).$$

Integrating the differential equation above, we have

$$\log\Phi = TF_1(\theta) - \int \theta\, d\theta/\lambda_1(\theta)$$
$$= TF_1(\theta) - \theta F_1(\theta) + \mathcal{G}_1(\theta) + C'(x').$$

where C' is a constant of integration, independent of θ, but possibly involving x'. Put

$$g_1(\theta) - \theta F_1(\theta) = n F_2(\theta),$$

whence

$$\Phi = \exp\{T F_1(\theta) + n F_2(\theta) + C'(x')\}.$$

Because Φ is a product of n similar factors φ
$\left[\text{i.e. } \Phi(x'|\theta) = \prod_{i=1}^{n} \varphi(x_i|\theta) \right]$

functions $f(x)$, $c(x)$ can be determined such that

$$T = \sum_{i=1}^{n} \frac{1}{n} f(x_i) \quad ; \quad C'(x') = \sum_{i=1}^{n} c(x_i).$$

The expression for Φ is consequently equivalent to

$$\varphi = \exp\{f(x) F_1(\theta) + F_2(\theta) + c(x)\} \tag{15}$$

For this distribution, then, $\frac{1}{n} \sum_i f(x_i)$ is an unbiassed estimate of θ, and has minimum sampling variance. The fact that φ obeys the total probability condition enables us, as we shall show in Section 3.6, to fix the hitherto arbitrary "constant of integration" $c(x)$.

3.5 Sufficient Statistics:- Equation (15) belongs to an important class of distributions which has been studied by B. O. Koopman, in his paper "Distributions Admitting a Sufficient Statistic" (Transactions of The American Mathematical Society, vol. 39 (1936) pp. 399-409). The concept of the "sufficiency" of a statistic was first formulated by R. A. Fisher, and the intuitive idea of such a statistic is that it is one which contains all the information given by the sample. Koopman asserts that this intuitive idea is expressed in the following definition.

Let $\varphi(x \mid \theta_1, \theta_2, \cdots \theta_\nu)$ be a probability distribution specified by ν parameters. The statistics $T_i(x_1, x_2 \cdots x_n)$ $(i=1,\cdots \nu)$ or $T_i(x')$ are <u>sufficient</u> if the equations

$$T_i(x') = T_i(y') \qquad (i = 1, 2, \cdots \nu)$$

imply the identity

$$\frac{\Phi(x' \mid \theta_1 \cdots \theta_\nu)}{\Phi(y' \mid \theta_1 \cdots \theta_\nu)} = \frac{\Phi(x' \mid \theta_1^x \cdots \theta_\nu^x)}{\Phi(y' \mid \theta_1^x \cdots \theta_\nu^x)}$$

where $\Phi(x' \mid \theta_1 \cdots \theta_\nu) = \prod_{i=1}^{n} \varphi(x_i \mid \theta_1 \cdots \theta_\nu)$.

The vectors x', y' denote any two possible samples of n observations, and $\theta_1 \cdots \theta_\nu$; $\theta_1^x \cdots \theta_\nu^x$ are any two possible sets of values of the parameters.

From this definition, it is possible to deduce the form of distributions which admit a system of sufficient statistics. Koopman gives a very general statement of the appropriate theorems, designed to cover those cases where φ has discontinuities. We are concerned only with continuous probability distributions, however, and the following enunciations will be adequate for our purpose.

<u>Koopman's First Theorem</u>:- $T_1(x'), T_2(x'), \cdots T_\nu(x')$ exist, and form a system of sufficient statistics for the distribution φ, if

$$\varphi(x \mid \theta_1 \cdots \theta_\nu) = \exp\left\{\sum_{k=1}^{\mu} f_k(x) F_k(\theta_1 \cdots \theta_\nu) + F'(\theta_1 \cdots \theta_\nu) + c(x)\right\}$$

where (i) F_k $(k = 1, 2, \cdots \mu)$ and F' are real, single-valued, analytic functions of $\theta_1, \theta_2, \cdots \theta_\nu$.

(ii) f_k $(k = 1, 2, \cdots \mu)$ are real, single-valued, analytic functions of x.

(iii) $\mu \leq \nu$. ($\mu = 0$ implies that all the f_k or F_k are lacking)

(iv) φ satisfies the total probability condition over the whole range of the variate x.

Further,

$$\sum_{i=1}^{n} f_k(x_i) = V_k(T_1, T_2, \cdots T_\nu) \qquad (k = 1, 2, \cdots \mu)$$

where V_k is a single-valued function of its ν arguments.

<u>Koopman's Second Theorem</u>:- Given the distribution

$$\varphi = \exp\left\{\sum_{k=1}^{\mu} f_k(x) F_k(\theta_1 \cdots \theta_\nu) + F'(\theta_1 \cdots \theta_\nu) + c(x)\right\}$$

(where the f's, F_k's and F' are as above); then a set of functions $T_i(x')$ $(i = 1, 2 \cdots \nu)$ form a system of sufficient statistics if

$$\sum_{i=1}^{n} f_k(x_i) = V_k(T_1, \cdots T_\nu) \qquad (k = 1, 2, \cdots \nu)$$

where V_k is a real, single-valued function of its ν arguments.

<u>Koopman's Third Theorem</u> -states if a distribution admits a system of sufficient statistics, the system will be obtained by the Maximum Likelihood method of estimation.

3.5.1 Properties of Koopman's Distributions:-

(a) Koopman's distribution

$$\varphi = \exp\left\{\sum_{k=1}^{\mu} f_k(x) F_k(\theta_1 \cdots \theta_v) + F'(\theta_1 \cdots \theta_v) + c(x)\right\}$$

is invariant under a transformation of the variate.

If, in fact, we transform the variate x by means of $x = u(y)$, the probability distribution of y is $\varphi_1(y \mid \theta_1 \cdots \theta_v)$ where

$$\varphi_1(y \mid \theta_1 \cdots \theta_v)\,dy$$
$$= \exp\left\{\sum_{k=1}^{\mu} f_k\{u(y)\} F_k(\theta_1 \cdots \theta_v) + F'(\theta_1 \cdots \theta_v) + c\{u(y)\}\right\} \frac{du}{dy}\,dy.$$

Writing

$$f_k\{u(y)\} = f_k'(y)$$
$$c\{u(y)\} + \log\frac{du}{dy} = c'(y)$$

we obtain

$$\varphi_1(y \mid \theta_1 \cdots \theta_v) = \exp\left\{\sum_{k=1}^{\mu} f_k'(y) F_k(\theta_1 \cdots \theta_v) + F'(\theta_1 \cdots \theta_v) + c'(y)\right\},$$

which is of Koopman's form. This form is invariant, therefore, under any transformation of the variate.

(b) Koopman's distribution is invariant under any transformation of the parametric coefficients. This property is obvious, on considering, say, the transformation

$$\theta_j = g_j(\theta_1' \cdot \theta_2' \cdots \theta_v') \qquad (j = 1, 2, \cdots v)$$

and on substituting in the distribution above.

It follows that we lose nothing in generality by always taking a particular form of Koopman's distribution, viz., that obtained by putting

$$F_k(\theta_1 \cdots \theta_\nu) = \psi_k \qquad (k = 1, 2, \cdots \mu)$$

(If $\mu < \nu$ we take $\theta_k = \psi_k$ for $k = \mu+1, \mu+2, \cdots \nu$)

Denote $F'(\theta_1 \cdots \theta_\nu)$, when expressed in terms of the ψ's, by $F(\psi_1 \cdots \psi_\nu)$. Koopman's distribution is then written in what we may call "canonical" form,

$$\varphi = \exp\left\{\sum_{k=1}^{\mu} f_k(x)\psi_k + F(\psi_1 \cdots \psi_\nu) + c(x)\right\}.$$

<u>3.5.2</u> <u>Non-Uniqueness of Sufficient Statistics</u>:- We are now in a position to recognise whether a given distribution admits~~x~~ sufficient statistics, and to ascertain whether any suggested set of estimates is sufficient. (For the latter, we apply the test of the original definition, or invoke Koopman's Second Theorem). But even though a particular set does satisfy the test of sufficiency, it is not the only set which does so. This is apparent from the Second Theorem, in which the V_k represent <u>any</u> real, single-valued functions. The criterion of sufficiency does not therefore provide an unambiguous solution to the problem of estimation. We are still faced with the queries - are all sufficient statistics equally "good"? If so - why? If not - which should be chosen?

<u>Example</u>: We illustrate the non-uniqueness of sufficient statistics by considering the normal distribution

$$\varphi = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left\{ -\frac{(x-m)^2}{2\sigma^2} \right\}$$

or $\varphi = \exp\left\{ -\frac{x^2}{2\sigma^2} + \frac{m}{\sigma^2}\cdot x - \left(\frac{m^2}{2\sigma^2} + \log\sigma\right) - \frac{1}{2}\log 2\pi \right\}.$

This is of Koopman's form (with $f_1 = -x^2$; $f_2 = x$; $F_1 = \frac{1}{2\sigma^2}$; $F_2 = \frac{m}{\sigma^2}$; $F = -\left(\frac{m^2}{2\sigma^2} + \log\sigma\right)$; $c = -\frac{1}{2}\log 2\pi$.

Accordingly, it admits a pair of sufficient statistics.

Are $T_1 = \frac{1}{n}\sum_i x_i = \bar{x}$; $T_2 = \frac{1}{n}\sum_i (x - \bar{x})^2$ sufficient?

Applying the test of Koopman's definition, we note that the identity

$$\Phi(x' \mid m, \sigma) / \Phi(y' \mid m, \sigma) \equiv \Phi(x' \mid m^*, \sigma^*) / \Phi(y' \mid m^*, \sigma^*)$$

is true if

$$\frac{(2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum(x-m)^2\right\}}{(2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum(y-m)^2\right\}} \equiv \frac{(2\pi)^{-\frac{n}{2}}\sigma^{*-n}\exp\left\{-\frac{1}{2\sigma^{*2}}\sum(x-m^*)^2\right\}}{(2\pi)^{-\frac{n}{2}}\sigma^{*-n}\exp\left\{-\frac{1}{2\sigma^{*2}}\sum(y-m^*)^2\right\}}$$

i.e., if (taking logs)

$$\frac{1}{2\sigma^2}\sum\left(x^2 - 2mx - y^2 + 2my\right) \equiv \frac{1}{2\sigma^{*2}}\sum\left(x^2 - 2m^*x - y^2 + 2m^*y\right)$$

i.e., if

$$\left(-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^{*2}}\right)\left\{\sum(x-\bar{x})^2 - \sum(y-\bar{y})^2 + \frac{\left(\sum x\right)^2 - \left(\sum y\right)^2}{n^2}\right\} + \left(\frac{m}{\sigma^2} - \frac{m^*}{\sigma^{*2}}\right)\left(\sum x - \sum y\right) \equiv 0$$

Now this is implied by $\bar{x} = \bar{y}$; $\frac{1}{n}\sum(x-\bar{x})^2 = \frac{1}{n}\sum(y-\bar{y})^2$

It is also
(as we see from the penultimate identity above(implied by

$$\bar{x} = \bar{y}\ ; \quad \frac{1}{n}\sum x^2 = \frac{1}{n}\sum y^2$$

By definition, therefore, $\bar{x} = \frac{1}{n}\sum x$ and $\frac{1}{n}\sum(x-\bar{x})^2$ are a pair of sufficient statistics for the normal distribution; while $\bar{x}$ and $\frac{1}{n}\sum x^2$ are a second such pair. We notice that $\bar{x}$ and $\frac{1}{n-1}\sum(x-\bar{x})^2$ are also sufficient. The decision which of these three pairs is the "best" for estimating purposes, would require the formulation of some criterion additional to that of sufficiency.

3.6 Koopman's Distributions and The Unbiassed Statistic of Minimum Variance:- Distributions which admit an unbiassed statistic of minimum variance are of the form (equation (15), Section 3.4)

$$\varphi = \exp\left\{ f(x)F_1(\theta) + F_2(\theta) + c(x) \right\}$$

and the statistic in question is $T = \sum_{i=1}^{n} \frac{1}{n} f(x_i)$ As this distribution is of Koopman's form, it admits a sufficient statistic. Moreover, by Koopman's Second Theorem, our unbiassed statistic T of minimum variance is sufficient.

In the foregoing expression $F_2(\theta)$ and $F_1(\theta)$ are not wholly independent. Let us therefore take the general Koopman distribution of one parameter

$$\varphi = \exp\left\{ \psi f(x) + F(\psi) + c(x) \right\}$$

and ascertain whether an unbiassed statistic of minimum variance can always be found.

Suppose such a statistic T exists for the estimation of $\theta(\psi)$. We have

$$\Phi = \exp\left\{\psi \sum_{i=1}^{n} f(x_i) + nF(\psi) + \sum_{i=1}^{n} c(x_i)\right\}$$

$$\therefore \frac{\partial \log \Phi}{\partial \theta} = \left\{\sum_i f(x_i) + n\frac{dF}{d\psi}\right\}\frac{d\psi}{d\theta}$$

Choose $\theta = -dF/d\psi$ so that this relation becomes

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \theta = -\left(\frac{d^2F}{n\,d\psi^2}\right)\frac{\partial \log \Phi}{\partial \theta}$$

which is of the form (9). Therefore $\frac{1}{n}\sum_i f(x_i)$ is an unbiassed estimate of $-dF/d\psi$, and has minimum variance. The actual value of the variance is (equation (14)) $-\frac{1}{n}\cdot\frac{d^2F}{d\psi^2}$.

3.6.1. Poissonian Sample:- When a sample of n observations of the general Koopman distribution of one parameter is Poissonian,

$$\Phi(x'|\theta') = \exp\left\{\sum_{i=1}^{n} \psi_i f(x_i) + \sum_{i=1}^{n} F(\psi_i) + \sum_{i=1}^{n} c(x_i)\right\}.$$

Hence $\left(f(x_i) + \frac{dF}{d\psi_i}\right)\frac{d\psi_i}{d\theta_i} = \frac{\partial \log \Phi(x'|\psi')}{\partial \theta_i}$

Choose $\theta_i = -dF(\psi_i)/d\psi_i$; $\bar{\theta} = \frac{1}{n}\sum_i \theta_i$. Then

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) - \bar{\theta} = -\sum_{i=1}^{n}\frac{1}{n}\frac{d^2F(\psi_i)}{d\psi_i^2}\frac{\partial \log \Phi(x'|\psi')}{\partial \theta_i}$$

This is of the form (10). Therefore $T = \frac{1}{n}\sum_i f(x_i)$ is an unbiassed estimate of $-\frac{1}{n}\sum_i \frac{dF(\psi_i)}{d\psi_i}$, and has minimum variance, of value (equation (12))

$$-\frac{1}{n^2}\sum_i \frac{d^2F(\psi_i)}{d\psi_i^2}$$

We note the result that, for Koopman's one parameter distribution, the unbiassed statistic of minimum variance (which always exists) is the same for both Bernouillian and Poissonian samples, viz., $\frac{1}{n} \sum_i f(x_i)$

3.6.2 Basic Parametric Coefficients:- An unbiassed statistic T of minimum variance (in Bernouillian samples) exists if

$$(T - \theta)\,\Phi = \lambda_1(\theta)\,\frac{\partial \Phi}{\partial \theta}$$

The Koopman distribution, we have seen, can be cast into this form by taking $\theta = -dF/d\psi$. It is, moreover, apparent from Section 3.6 that (apart from multiples of θ) no other function of ψ exists which leads to this requisite form. This particular function of the parameter ψ will be termed the "basic parametric coefficient." We may define it formally as a coefficient which permits of estimation by an unbiassed statistic of minimum variance.

The theory of estimation by Moments, and the theory of Maximum Likelihood, both fail to indicate the phenomenon of the basic parametric coefficient. While these methods of estimation lead to the "best" statistic - according to their own standards - for a prescribed coefficient, neither of them tells which coefficient we should seek to estimate. They do not suggest, either, whether or not it is "desirable" to estimate one parameter rather than some function thereof. The method of "unbiassed minimum variance," which leads not only to a particular statistic, but also to a particular coefficient, thus sheds light on a new, perhaps unsuspected, aspect of the theory of estimation.

In Section 3.5.2, we showed how the criterion of sufficiency failed to solve unambiguously the problem of estimation. By adopting the properties of unbiassedness and of minimum variance as criteria, we obviate this difficulty; for these postulates provide a rational basis on which to make a unique selection from among the possible sufficient statistics.

3.7 The Total Probability Condition and The Unbiassed Statistic of Minimum Variance:- Given T (the unbiassed statistic of minimum variance) and $\lambda_1(\theta)$, integration of the equation

$$(T-\theta)\,\Phi = \lambda_1(\theta)\,\frac{\partial \Phi}{\partial \theta}$$

leads, as in Section 3.4, to

$$\varphi = \exp\{\psi f(x) + F(\psi) + c(x)\}$$

where $T = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$ and where F is known in terms of λ_1. The term $c(x)$ is, however, arbitrary, and arises as a constant in the integration with respect to θ. If φ represents an actual probability distribution, it obeys the total probability condition, viz., $\int_a^b \varphi\, dx = 1$ for all values of ψ (or θ), where a, b are the limits of the variate x. The questions may now be posed - does this condition serve to determine the hitherto arbitrary $c(x)$?

Does it determine $c(x)$ uniquely? Is it possible to specify T and $\lambda_1(\theta)$ in such a way that no function $c(x)$ can be determined?

Consider, in the first instance, that the range of x is $(-\infty, \infty)$. We wish to determine $c(x)$ such that

$$\int_{-\infty}^{\infty} exp\{\psi f(x) + F(\psi) + c(x)\}\, dx = 1$$

for all values of ψ; or writing $exp\, c(x) = C_1(x)$,

$$\int_{-\infty}^{\infty} C_1(x)\, e^{\psi f(x)}\, dx = e^{-F(\psi)}$$

Now introduce a change of variable, and put $f(x) = y$. <u>Assume</u> that the limits of y remain $(-\infty, +\infty)$ and let $x = f^{-1}(y)$ denote the inverse transformation. We have

$$\int_{-\infty}^{\infty} C_1\{f^{-1}(y)\}\, e^{\psi y}\, \frac{df^{-1}(y)}{dy}\, dy = e^{-F(\psi)}$$

or, writing $C_1\{f^{-1}(y)\}(df^{-1}(y)/dy) = C(y)$,

$$\int_{-\infty}^{\infty} C(y)\, e^{\psi y}\, dy = e^{-F(\psi)},$$

which holds for all ψ. We may therefore replace ψ by $-\psi$, and our equation becomes

$$\int_{-\infty}^{\infty} C(y)\, e^{-\psi y}\, dy = e^{-F(-\psi)}$$

Thus our unknown $C(y)$ has a known Laplace transform, viz., $\{exp - F(-\psi)\}$. At least formally, we may invert by Mellin's Theorem, whereupon we obtain

$$C(y) = \frac{1}{2\pi i}\int_{\alpha - i\infty}^{\alpha + i\infty} \exp\{zy - F(-z)\}\, dz.$$

α is any real number lying to the right of all singularities of $\exp\{-F-(z)\}$ in the complex z-plane. By the general theory of Laplace transforms, $C(y)$ is unique. Its validity in any particular case can be tested either by the relevant theorems in transform theory, or by actual substitution in $\int_{-\infty}^{\infty} C(y)\, e^{-\psi y}\, dy$.

When the limits of y are $(0, \infty)$ the function $C(y)$, and hence the original $c(x)$, can again be found uniquely by means of a Mellin transform. When the limits of y (or of x) are finite - say (a, b) - we have the linear integral equation $\int_a^b C_1(x)\, e^{\psi f(x)}\, dx = e^{-F(\psi)}$.
This can always be solved for $C_1(x)$ by the method of Chapter Two, the solution so obtained being unique of class L^2.

Hence (i) if $f(x)$, $F(\psi)$, and the range of x, are given, we can determine uniquely the remaining element $c(x)$ of the Koopman distribution;

or, an alternative statement of the same result,

(ii) there is a unique distribution, over a given range, for which a prescribed function T is an unbiassed statistic of minimum variance of prescribed magnitude λ_1.

(The range must of course be independent of the parameter, a stipulation made at the beginning of this chapter)

3.7.1 Examples:-

(a) In the equation

$$T - \theta = \lambda_1(\theta)\, \frac{\partial \log \Phi}{\partial \theta}$$

let us choose the variance $\lambda_1(\theta)$ as independent of θ, and equal to σ^2/n say, where σ is a real constant. Choose for T the mean of the sample, $\frac{1}{n}\sum_{i=1}^{n} x_i$ Integration of the equation gives

$$\log \Phi = -\frac{n\theta^2}{2\sigma^2} + \frac{\theta}{\sigma^2}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} c(x_i)$$

where the last term represents a constant of integration, independent of θ but possibly involving the x's. Hence

$$\log \varphi = -\frac{\theta^2}{2\sigma^2} + \frac{\theta x}{\sigma^2} + c(x)$$

or

$$\varphi = \exp\left\{-\frac{\theta^2}{2\sigma^2} + \frac{\theta x}{\sigma^2} + c(x)\right\}.$$

Let us furthermore fix the range of x as $(-\infty, +\infty)$. We must now determine $c(x)$ so that

$$\int_{-\infty}^{\infty} \varphi\, dx = 1 \quad \text{for all values of } \theta.$$

Writing $\exp c(x) = C(x)$, this equation may be rewritten as

$$\int_{-\infty}^{\infty} C(x)\, e^{\theta x/\sigma^2}\, dx = e^{\theta^2/2\sigma^2}$$

or (putting $\theta/\sigma^2 = -\psi$)

$$\int_{-\infty}^{\infty} C(x)\, e^{-\psi x}\, dx = e^{\psi^2\sigma^2/2}.$$

Inverting by Mellin's Theorem,

$$C(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \exp\left\{xz + \frac{z^2\sigma^2}{2}\right\} dz$$

or (writing $z = a + iy$)

$$C(x) = \frac{1}{2\pi} e^{-\frac{x^2}{2\sigma^2}} \int_{-\infty}^{\infty} \exp\left\{-\left[y - \frac{i(x+\sigma^2 a)}{\sigma^2}\right]^2 \cdot \frac{\sigma^2}{2}\right\}$$

We evaluate this integral as follows: Consider the contour integral $\int_\Gamma e^{-\lambda z^2} dz$, where λ is real and > 0, and where Γ denotes the rectangle (in the Argand diagram) with vertices $-R,\ R,\ -R+i\alpha,\ R+i\alpha\quad [R, \alpha \text{ real}]$.

Along the line $-R, R$ the contribution to the integral is

$$\int_{-R}^{R} e^{-\lambda x^2}\, dx.$$

Along the line $R+i\alpha, -R+i\alpha$, the contribution is

$$\int_{R}^{-R} e^{-\lambda(x+i\alpha)^2}\, dx.$$

On the line $R, R+i\alpha$ the absolute value of the contribution is

$$\left|\int_0^\alpha e^{-\lambda(R+iy)^2} dy\right| = e^{-\lambda R^2}\left|\int_0^\alpha e^{\lambda y^2 - 2i\lambda R y} dy\right|$$

$$\leq e^{-\lambda R^2}\int_0^\alpha |e^{\lambda y^2}|\, dy \leq \alpha e^{-\lambda R^2 + \lambda\alpha^2}$$

$\to 0$, as $R \to \infty$, for all finite values of α, since $\lambda > 0$. Similarly the contribution from the final limb of the rectangle tends to zero as $R \to \infty$.

The integrand has no singularities within the contour. In the limit, therefore, as $R \to \infty$,

$$\int_{-\infty}^{\infty} e^{-\lambda x^2} dx + \int_{\infty}^{-\infty} e^{-\lambda(x+i\alpha)^2} dx = 0 .$$

Now choose $\lambda = \sigma^2/2$ (which is > 0) and $\alpha = -\dfrac{x+\sigma^2 a}{\sigma^2}$ (which is real); then

$$\int_{-\infty}^{\infty} exp\left\{-\left[y - \frac{i(x+\sigma^2 a)}{\sigma^2}\right]^2 \cdot \frac{\sigma^2}{2}\right\} dy = \int_{-\infty}^{\infty} e^{-\frac{\sigma^2 x^2}{2}} dx = \frac{\sqrt{2\pi}}{\sigma}.$$

Returning to our last expression for $C(x)$, we insert this value of the integral, to obtain

$$C(x) = \frac{1}{\sqrt{2\pi}\,\sigma} exp\left\{-\frac{x^2}{2\sigma^2}\right\}.$$ Consequently

$$\varphi = C(x)\, exp\left\{-\frac{\theta^2}{2\sigma^2} + \frac{\theta x}{\sigma^2}\right\}$$
$$= \frac{1}{\sqrt{2\pi}\cdot\sigma} exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$$

As we know that this satisfies the total probability condition, the formal use of Mellin's Theorem is justified. Thus, for the normal distribution (with given variance σ^2) the <u>mean</u> is an unbiassed statistic of minimum variance. The mean is therefore, by Section 3.6, sufficient, and (remembering that the function $C(x)$ was unique) we state the following:

<u>Theorem</u>:- The only one-parameter distribution, over the range $(-\infty, +\infty)$, for which the mean of the sample is a sufficient statistic, with sampling variance independent

of the parameter estimated, is the normal distribution. The two-parameter analogue of this theorem has been given as "the only Pearsonian distribution for which the mean is a sufficient statistic is the normal distribution." We shall discuss this later, in Chapter Five, but we may here note that, for the one-parametric case at any rate

(i) The adjective "Perasonian" may be removed.

(ii) the proviso regarding the sampling variance of the mean is necessary. This necessity is further illustrated in the next example.

Example (b):- In the equation

$$T - \theta = \lambda_1(\theta) \frac{\partial \log \Phi}{\partial \theta}$$

let us choose the variance $\lambda_1(\theta)$ as θ^2/n. Take T again as the mean of the sample. Integrating,

$$\log \Phi = -\frac{1}{\theta}\sum_{i=1}^{n} x_i - n \log \theta + \sum_{i=1}^{n} c(x_i)$$

where c is the constant of integration.

Hence
$$\varphi = \frac{1}{\theta} \exp\left\{-\frac{x}{\theta} + c(x)\right\}.$$

Choose the range of the variate this time as $(0, \infty)$, so that

$$\int_0^\infty \varphi \, dx = 1 \quad \text{for all} \quad \theta.$$

Putting $C(x) = \exp c(x)$, we have

$$\int_0^\infty C(x) e^{-x/\theta} dx = \theta$$

or (writing $\lambda = 1/\theta$), $\int_0^\infty C(x) e^{-\lambda x} dx = 1/\lambda.$

By Mellin's Theorem,

$$C(x) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \frac{e^{zx}}{z}\,dz \quad (a>0).$$

Considering this integral, we note that, when $\frac{\pi}{2} \leq \arg z \leq \frac{3\pi}{2}$, the integrand tends to zero, as $|z| \to \infty$, uniformly with respect to $\arg z$. So the integral is equal to

$$I = \int_\Gamma \frac{e^{zx}}{z}\,dz$$

where Γ denotes the contour formed by the line $a-i\infty$, $a+i\infty$ $(a>0)$, and the infinite semi-circle in the third and fourth quadrants. The only singularity within Γ is at $z=0$, which is obviously a simple pole, where the residue is

$$\lim_{z\to 0} e^{zx} = 1$$

Hence

$$I = 2\pi i, \quad \text{or}$$
$$C(x) = 1.$$

Therefore $\varphi = \frac{1}{\theta}e^{-x/\theta}$, which actually satisfies the total probability condition over $(0,\infty)$ provided θ is positive. This condition is fulfilled, since θ is the mean value, over all possible samples of n, of $\frac{1}{n}\sum_{i=1}^{n} x_i$, and $x \geqslant 0$.

Thus, in the distribution

$$\frac{1}{\theta}e^{-x/\theta} \qquad (x \geqslant 0)$$

θ is estimated without bias by $\frac{1}{n}\sum_i x_i$, which statistic has minimum variance, of value θ^2/n. The mean is consequently a sufficient statistic.

3.8 Comparison With Maximum Likelihood:- Let us tackle by Maximum Likelihood the problem of estimation for the one-parameter Koopman distribution. The likelihood function, corresponding to

$$\varphi = \exp\{\psi f(x) + F(\psi) + c(x)\}$$

is, in a sample of n,

$$L = \psi \sum_i f(x_i) + nF(\psi) + \sum_i c(x_i).$$

To estimate a coefficient θ (i.e., some function of ψ) the precept is to equate $\frac{\partial L}{\partial \theta}$ to zero. The solution of this equation gives the required estimate, $\hat{\theta}$ say. Since

$$\frac{\partial L}{\partial \theta} = \left\{ \sum_i f(x_i) + n\frac{dF}{d\psi} \right\} \frac{d\psi}{d\theta}$$

the rule $\frac{\partial L}{\partial \theta} = 0$ gives

$$\sum_i f(x_i) + n\frac{dF}{d\psi} = 0,$$

as $\frac{d\psi}{d\theta} \neq 0$ (i.e., since θ is not a mere constant). Therefore

$$-\widehat{\left(\frac{dF}{d\psi}\right)} = \frac{1}{n} \sum_i f(x_i).$$

That is, $\frac{1}{n} \sum_i f(x_i)$ is an estimate of $-dF/d\psi$. To follow out our original purpose, we solve this last equation so as to obtain, not $-\widehat{(dF/d\psi)}$ but θ, in terms of the f's. Of course, if θ is chosen as $-dF/d\psi$, we arrive at exactly the same statistic as was yielded by the criteria of unbiassedness and minimum vaŕiance.

We notice, too, that the method of Maximum Likelihood leads to the equation

$$\sum_i f(x_i) + n\frac{dF}{d\psi} = 0,$$

whatever coefficient θ we are estimating for - a result which is geometrically obvious. This does <u>not</u> mean that Maximum Likelihood leads to a basic parametric coefficient, since this theory gives no indication that $-dF/d\psi$ should be estimated, rather than ψ or any other function of ψ.

That the two methods of estimation yield the same statistic for the Koopman distribution is not surprising. We have seen that the unbiassed statistic of minimum variance is sufficient; and we recall (Koopman's Third Theorem, Section 3.5) that if a sufficient statistic exists, it is found by the method of Maximum Likelihood.

<u>3.8.1.</u> <u>Variance</u>:- One of the striking results of Likelihood theory is the simple expression, valid in indefinitely large samples, for the variance of the estimates. Denoting by $\sigma^2_{\hat{\theta}}$ the sampling variance of the estimate $\hat{\theta}$, the formula in question is

$$-\frac{1}{\sigma^2_{\hat{\theta}}} = n\bar{b}$$

where

$$\bar{b} = \int \varphi \frac{\partial^2 \log \varphi}{\partial \theta^2}\, dx$$

Applying this to the calculation of the variance of $\frac{1}{n}\sum_i f(x_i)$ in the Koopman distribution, we have

$$\log \varphi = \psi f(x) + F(\psi) + c(x)$$

$$\frac{\partial^2 \log \varphi}{\partial \theta^2} = \left(f(x) + \frac{dF}{d\psi}\right)\frac{d^2\psi}{d\theta^2} + \frac{d^2F}{d\psi^2}\left(\frac{d\psi}{d\theta}\right)^2$$

so

$$\bar{b} = \frac{d^2\psi}{d\theta^2}\int\left(f(x) + \frac{dF}{d\psi}\right)\varphi\, dx + \frac{d^2F}{d\psi^2}\left(\frac{d\psi}{d\theta}\right)^2\int \varphi\, dx .$$

Now $\int \varphi\, dx = \int \exp\{\psi f(x) + F(\psi) + c(x)\}\, dx = 1$, for all values of ψ. Differentiating with respect to ψ,

$$\int\left(f(x) + \frac{dF}{d\psi}\right)\varphi\, dx = 1$$

Accordingly

$$\bar{b} = \frac{d^2F}{d\psi^2}\left(\frac{d\psi}{d\theta}\right)^2$$

Since $\frac{1}{n}\sum_i f(x_i)$ estimates $\theta = -dF/d\psi$,

$$\frac{d\psi}{d\theta} = -\left(\frac{d^2F}{d\psi^2}\right)^{-1} \quad \text{and} \quad \bar{b} = \left(\frac{d^2F}{d\psi^2}\right)^{-1} .$$

Substitution in the formula

$$\sigma^2_{\hat\theta} = -1/n\bar{b}$$

gives

$$\sigma^2_{\hat\theta} = -\frac{1}{n}\frac{d^2F}{d\psi^2}$$

for the sampling variance of $\frac{1}{n}\sum_i f(x_i)$. The general theory of Likelihood asserts merely that this value is valid in indefinitely large samples. In point of fact, it is correct whatever the size of the sample (Section 3.6)

This last remark draws attention to a feature worthy of emphasis - that all our results thus far hold irrespective of the number of observations in the sample, be they few or many. Furthermore, we have made no assumptions whatsoever regarding the distribution of statistics, whether in large or small samples. Our conclusions have been founded on the two initial postulates - that a statistic should be unbiassed and should have minimum variance - without qualification.

Chapter Four

The Estimation of Two Parameters by Unbiassed Statistics of Minimum Variance.

4.0 The problem of the simultaneous estimation of several parameters not unnaturally presents greater difficulty than does the corresponding case of only one coefficient. Trouble arises most acutely, perhaps, on a practical level. It is comparatively straightforward to formulate criteria by which statistics may be judged; but, given some particular distribution, it is often impossible to obtain any estimates at all which satisfy the chosen tests. Our criteria, in short, tend to become too exclusive.

We shall meet this especial difficulty in due course. Meanwhile, let us proceed to the estimation of the two coefficients in the distribution $\varphi(x|\theta_1 \cdot \theta_2)$ by means of statistics which shall

(i) be symmetrical functions of the n observations of the sample.

(ii) be unbiassed estimates of θ_1 and θ_2 respectively

(iii) have respectively a small sampling variance than any other unbiassed estimate of θ_1, or of θ_2.

The usual restrictions will be imposed on $\varphi(x/\theta_1 \cdot \theta_2)$; that is, it is assumed to be continuous in all three variables, and to have continuous partial derivatives of all orders. The range of the variate x will be taken as independent of the coefficients θ_1, θ_2; it may, however, be either finite or infinite.

Our requirements are expressed symbolically as :

$$\int (T_1 - \theta_1)\, \Phi \, dx' = 0 \qquad ; \int \left(T_2 - \theta_2\right) \Phi \, dx' = 0.$$

$$V_{11} = \int (T_1 - \theta_1)^2 \Phi \, dx' = \text{minimum} \;; V_{22} = \int (T_2 - \theta_2)^2 \Phi \, dx' = \text{minimum}.$$

where

$$\Phi = \prod_{i=1}^{n} \varphi(x_i \mid \theta_1 \cdot \theta_2) \tag{1}$$

The Total probability condition is

$\int \varphi \, dx = 1$ for all values of θ_1 and of θ_2,

whence $\int \Phi \, dx' = 1$ for all values of θ_1 and of θ_2.

T_1, T_2 are found by a method identical with that developed in Chapter Three. Thus, to consider the case of T_1, in detail: let there exist some or all of the set of quantities

$$\lambda_{11},\ \lambda_{12};\ \lambda_{1(1\cdot1)},\ \lambda_{1(1\cdot2)},\ \lambda_{1(2\cdot2)};\ \cdots\cdots$$

$$\mu_{1(1\cdot1)},\ \mu_{1(1\cdot2)},\ \mu_{1(2\cdot2)};\ \mu_{1(1\cdot1\cdot1)},\ \mu_{1(1\cdot1\cdot2)},\ \mu_{1(2\cdot2\cdot1)},\ \mu_{1(2\cdot2\cdot2)};\ \cdots\cdots$$

[which are functions of θ_1, θ_2 but which do not involve x']

such that

$$T_1 \equiv \theta_1 + \frac{1}{\Phi}\left(\lambda_{11}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{12}\frac{\partial \Phi}{\partial \theta_2}\right) + \frac{1}{\Phi}\left(\lambda_{1(1\cdot1)}\frac{\partial^2 \Phi}{\partial \theta_1^2} + 2\lambda_{1(1\cdot2)}\frac{\partial^2 \Phi}{\partial \theta_1 \partial \theta_2} + \lambda_{1(2\cdot2)}\frac{\partial^2 \Phi}{\partial \theta_2^2}\right.$$

$$+ \cdots\cdots$$

$$+ \sum_{i=1}^{n}\left\{\frac{1}{\varphi(x_i \mid \theta_1 \cdot \theta_2)}\left(\mu_{1(1\cdot1)}\frac{\partial^2}{\partial \theta_1^2} + 2\mu_{1(1\cdot2)}\frac{\partial^2}{\partial \theta_1 \partial \theta_2} + \mu_{1(2\cdot2)}\frac{\partial^2}{\partial \theta_2^2}\right)\varphi(x_i \mid \theta_1 \cdot \theta_2)\right\}$$

$$+ \sum_{i=1}^{n}\left\{\frac{1}{\varphi(x_i \mid \theta_1 \cdot \theta_2)}\left(\mu_{1(1\cdot1\cdot1)}\frac{\partial^3}{\partial \theta_1^3} + 3\mu_{1(1\cdot1\cdot2)}\frac{\partial^3}{\partial \theta_1^2 \partial \theta_2} + 3\mu_{1(2\cdot2\cdot1)}\frac{\partial^3}{\partial \theta_1 \partial \theta_2^2}\right.\right.$$

$$\left.\left. + \mu_{1(2\cdot2\cdot2)}\frac{\partial^3}{\partial \theta_2^3}\right)\varphi(x_i \mid \theta_1 \cdot \theta_2)\right\} + \cdots\cdots \tag{2}$$

depends on x' alone, and not on θ_1, θ_2. Such a function

T_1 is symmetric in its n arguments x', and it satisfies the conditions of unbiassedness

$$\int (T_1 - \theta_1) \Phi \, dx' = 0 .$$

This last statement is a consequence of the total probability condition which implies

$$\int \frac{\partial^{r+s} \Phi}{\partial \theta_1^r \partial \theta_2^s} dx' = 0$$

$$\int \frac{1}{\varphi(x_i | \theta_1 \cdot \theta_2)} \frac{\partial^{r+s} \varphi(x_i|\theta_1 \cdot \theta_2)}{\partial \theta_1^r \partial \theta_2^s} \Phi \, dx' = \int \frac{\partial^{r+s} \varphi(x_i | \theta_1 \cdot \theta_2) dx_i}{\partial \theta_1^r \partial \theta_2^s} \left\{ \int \varphi \, dx \right\}^{n-1}$$

$= 0$ for all values of θ_1, θ_2.

$$\left[\begin{matrix} r = 0, 1, 2 \cdots \\ s = 0, 1, 2 \cdots \end{matrix} \right\} r+s \neq 0 \Bigg]$$

Differentiation under the integral sign is legimate since φ, and its derivatives, are by hypothesis continuous functions of all their variables.

φ obeys the multiplicative law of independent probabilities expressed by equation (1). Repeating the analysis of Section 3.0.1, we find that the necessary and sufficient condition for the compatibility of (1) and (2) is

$$\lambda_{1\,(1\cdot 1)} = \lambda_{1\,(1\cdot 2)} = \lambda_{1\,(2\cdot 2)} = \cdots\cdots = 0 \qquad (3)$$

Suppose next, that T_1, is changed by a small arbitrary amount to $T_1 + \epsilon_1 f_1(x')$. The variance is correspondingly altered to

$$V_{11} + \delta V_{11} = \int \left(T_1 + \epsilon_1 f_1(x') - \theta_1 \right)^2 \Phi \, dx' .$$

When ϵ_1 is so small that terms in ϵ_1^2 are negligible, the first variation of V_{11} is

$$\delta V_{11} = 2\epsilon_1 \int (T_1 - \theta_1) f_1(x') \Phi \, dx' .$$

In order that T_1 should make the variance stationary, we require $\delta V_{11} = 0$. As in Chapter Three, it is obviously necessary to restrict the choice of the arbitrary $f_1(x')$. We accordingly stipulate that $f_1(x')$ must be of the form

$$f_1(x') = \sum_r U_r + \sum_s b_s u_s$$

where (4)

(i) $$U_r = \sum_{j=0}^{r} \frac{a_{r-j \cdot j}}{\Phi} \cdot \frac{\partial^r \Phi}{\partial \theta_1^{r-j} \partial \theta_2^{j}}$$

U_r is to be evaluated for the actual values of θ_1, θ_2 which specify the population. The terms $a_{r-j \cdot j}$ are arbitrary numerical constants . Thus U_r is a function of x' alone

(ii) u_s is any solution of the n-fold linear integral equation

$$\int u(x') \Phi(x' | \theta_1 \cdot \theta_2) \, dx' = h_s(\theta_2) ,$$

the right hand side denoting any arbitrary continuous function of θ_2 . This equation has an infinite number of solutions, provided $n > 2$, and we have seen how to obtain as many as may be required.

(iii) the b_s are arbitrary numerical constants

(iv) the summations $\sum_r , \sum_s$ cover as many terms as one pleases .

Substitute $f_1(x')$ as given by (4), and T_1 , as given by (2) and (3), into the expression for δV_{11} . Clearly a necessary condition for $\delta V_{11} = 0$ is

$$\mu_{1(1\cdot1)} = \mu_{1(1\cdot2)} = \mu_{1(2\cdot2)} = \mu_{1(1\cdot1\cdot1)} = \cdots = 0 .$$

When these functions are zero, we are left with

$$\delta V_{11} = 2\epsilon_1 \int \left(\lambda_{11} \frac{\partial \Phi}{\partial \theta_1} + \lambda_{12} \frac{\partial \Phi}{\partial \theta_2} \right) f_1(x')\, dx'$$

$$= 2\epsilon_1 \lambda_{11} \frac{\partial}{\partial \theta_1} \int f_1(x') \Phi\, dx' + 2\epsilon_1 \lambda_{12} \frac{\partial}{\partial \theta_2} \int f_1(x') \Phi\, dx'.$$

Now $\int f_1(x') \Phi\, dx' = \sum_r \sum_j a_{r-j\cdot j} \int \frac{\partial^r \Phi}{\partial \theta_1^{r-j} \partial \theta_2^j}\, dx' + \sum_s b_s \int u_s \Phi\, dx'$

$$= \sum_s b_s h_s(\theta_2),$$

since each term in the first (double) sum is zero.

Therefore

$$\frac{\partial}{\partial \theta_1} \int f_1(x') \Phi\, dx' = 0 \quad ; \quad \frac{\partial}{\partial \theta_2} \int f_1(x') \Phi\, dx' = \sum_s b_s \frac{dh_s(\theta_2)}{d\theta_2}$$

and

$$\delta V_{11} = 2\epsilon_1 \lambda_{12} \sum_s b_s \frac{dh_s(\theta_2)}{d\theta_2} .$$

Since the h_s are a set of arbitrary functions of θ_2, the necessary and sufficient condition for $\delta V_{11} = 0$ is

$$\lambda_{12} = 0 .$$

To sum up - if there exists a function λ_{11} [a function of θ_1 , θ_2 , but not of x'] such that

$$T_1 \equiv \theta_1 + \frac{\lambda_{11}}{\Phi} \frac{\partial \Phi}{\partial \theta_1} \qquad (5)$$

depends on x' alone, then T_1 is an unbiassed estimate of θ_1, and makes the variance, for small variations of the type permitted by (4), stationary.

From the change in variance when T_1 is changed to $T_1 + \epsilon_1 f_1(x')$ viz

$$\int (T_1 + \epsilon_1 f_1(x') - \theta_1)^2 \Phi \, dx' - \int (T_1 - \theta_1)^2 \Phi \, dx',$$

we obtain the second variation, or the term in ϵ^2. It is

$$\delta_2 V_{11} = \epsilon_1^2 \int f_1^2 \Phi \, dx' > 0$$

since Φ, the probability distribution, is > 0 everywhere. Consequently T_1 makes the variance a <u>minimum</u>.

Similarly, it can be shown that if a function $\lambda_{22}(\theta_1 \cdot \theta_2)$ exists, such that

$$T_2 \equiv \theta_2 + \frac{\lambda_{22}}{\Phi} \cdot \frac{\partial \Phi}{\partial \theta_2} \qquad (6)$$

depends on x' alone, then T_2 is an unbiassed estimate of θ_2, and has minimum variance. This conclusion is reached by comparing the variance of T_2 with that of $T_2 + \epsilon_2 f_2(x')$, where ϵ_2 is small and where $f_2(x')$ is an arbitrary function of the form

$$f_2(x') = \sum_r u_r + \sum_s b_s v_s .$$

u_r is as before, and v_s is any solution of the n-fold integral equation

$$\int v(x') \, \Phi(x' / \theta_1 \cdot \theta_2) \, dx' = h_s'(\theta_1) .$$

h_s' denotes any continuous function of θ_1 .

<u>4.0.1.</u> <u>Uniqueness</u>. The question whether T_1 , T_2 are unique can be investigated precisely along the lines developed in Sections 3.1 and 3.2. We find that they are not unique in Bernouillian samples. In Poissonian samples, however, there is no other statistic which has the same variance V_{11} as has T_1, and no other statistic has the same variance V_{22} as has T_2 . [V_{11} and V_{22} here denote functions of $2n$ variables, where n is the number in the sample.]

The uniqueness property justifies the absence of a rigorous proof that (2) represents the most general form of an unbiassed estimate of θ_1 .

<u>4.1</u> <u>Variances of the Unbiassed Statistics of Minimum Variance</u> :-

Let T_1, T_2 be unbiassed estimates, of θ_1 and θ_2 respectively, of minimum variance, so that

$$T_1 - \theta_1 = \lambda_{11} \frac{\partial \log \Phi}{\partial \theta_1}$$

$$T_2 - \theta_2 = \lambda_{22} \frac{\partial \log \Phi}{\partial \theta_2}$$

where λ_{11}, λ_{22} are each functions of θ_1 and θ_2.

The sampling variance of T_1 is

$$V_{11} = \int (T_1 - \theta_1)^2 \, \Phi \, dx' = \lambda_{11}^2 \int \left(\frac{\partial \log \Phi}{\partial \theta_1}\right)^2 \Phi \, dx'.$$

Now

$$\frac{\partial^2 \log \Phi}{\partial \theta_1^2} = \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial \theta_1^2} - \left(\frac{\partial \log \Phi}{\partial \theta_1}\right)^2$$

and, by virtue of the total probability condition,

$$\int \frac{\partial^2 \Phi}{\partial \theta_1^2} \, dx' = 0$$

Consequently

$$V_{11} = -\lambda_{11}^2 \int \frac{\partial^2 \log \Phi}{\partial \theta_1^2} \, \Phi \, dx' \qquad (7)$$

Another expression for the variance is obtained on differentiating the equation

$$\frac{\partial \log \Phi}{\partial \theta_1} = \frac{T_1 - \theta_1}{\lambda_{11}}$$

with respect to θ_1, which gives

$$\frac{\partial^2 \log \Phi}{\partial \theta_1^2} = -\frac{1}{\lambda_{11}} + (T_1 - \theta_1) \frac{\partial}{\partial \theta_1}\left(\frac{1}{\lambda_{11}}\right).$$

Multiply by Φ and integrate with respect to the n variates x'.

There results

$$\int \frac{\partial^2 \log \Phi}{\partial \theta_1^2} \Phi \, dx' = -1/\lambda_{11} + \frac{\partial}{\partial \theta_1}(1/\lambda_{11}) \int (T_1 - \theta_1) \Phi \, dx'$$

$$= -1/\lambda_{11},$$

since Φ satisfies the total probability condition, and since T_1 is an unbiassed estimate of θ_1.

Substitution in (7) yields, finally,

$$V_{11} = \lambda_{11} \tag{8}$$

In like manner, the variance of T_2 is

$$V_{22} = -\lambda_{22}^2 \int \frac{\partial^2 \log \Phi}{\partial \theta_2^2} \Phi \, dx' \tag{9}$$

or $$V_{22} = \lambda_{22} . \tag{10}$$

<u>4.1.1</u> Covariance of T_1, T_2:- The covariance of T_1 and T_2 is

$$V_{12} = \int (T_1 - \theta_1)(T_2 - \theta_2) \Phi \, dx'$$

$$= \lambda_{11} \lambda_{22} \int \frac{\partial \log \Phi}{\partial \theta_1} \cdot \frac{\partial \log \Phi}{\partial \theta_2} \Phi \, dx'$$

Now $$\frac{\partial^2 \log \Phi}{\partial \theta_1 \partial \theta_2} = \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial \theta_1 \partial \theta_2} - \frac{\partial \log \Phi}{\partial \theta_1} \cdot \frac{\partial \log \Phi}{\partial \theta_2}$$

Hence $$V_{12} = -\lambda_{11} \lambda_{22} \int \frac{\partial^2 \log \Phi}{\partial \theta_1 \partial \theta_2} \Phi \, dx' .$$

To simplify, differentiate

$$\frac{\partial \log \Phi}{\partial \theta_2} = \frac{T_2 - \theta_2}{\lambda_{22}}$$

with respect to θ_1, obtaining

$$\frac{\partial^2 \log \Phi}{\partial \theta_1 \partial \theta_2} = (T_2 - \theta_2) \frac{\partial}{\partial \theta_1} \left(\frac{1}{\lambda_{22}} \right) .$$

Multiplying by Φ and integrating with respect to the n variables x',

$$\int \frac{\partial^2 \log \Phi}{\partial \theta_1 \partial \theta_2} \Phi \, dx' = \frac{\partial}{\partial \theta_1} \left(\frac{1}{\lambda_{22}} \right) \int (T_2 - \theta_2) \Phi \, dx' = 0$$

Accordingly $$V_{12} = 0 .$$

That is, T_1 and T_2 are uncorrelated.

<u>4.1.2</u> <u>A Property of The Functions</u> $\lambda_{11}, \lambda_{22}$:- We have

seen that

$$\frac{\partial \log \Phi}{\partial \theta_1} = \frac{T_1 - \theta_1}{\lambda_{11}} \quad ; \quad \frac{\partial \log \Phi}{\partial \theta_2} = \frac{T_2 - \theta_2}{\lambda_{22}}$$

Hence

$$(T_2 - \theta_2)\frac{\partial \lambda_{22}^{-1}}{\partial \theta_1} = (T_1 - \theta_1)\frac{\partial \lambda_{11}^{-1}}{\partial \theta_2}$$

which implies either

(A) $$\frac{\partial \lambda_{22}^{-1}}{\partial \theta_1} = \frac{\partial \lambda_{11}^{-1}}{\partial \theta_2} = 0$$

or (B) if these quantities are not zero

$$\frac{T_2 - \theta_2}{T_1 - \theta_1} = \left(\frac{\partial \lambda_{11}^{-1}}{\partial \theta_2}\right) \Big/ \left(\frac{\partial \lambda_{22}^{-1}}{\partial \theta_1}\right)$$

= function of θ_1 and θ_2, independent of x'. Alternative (B) is unacceptable, because T_1, T_2 are functions of the n random observations $x' = (x_1, x_2, \cdots x_n)$. Therefore alternative (A) must hold. That is, λ_{11} must be a function, not of θ_1 and θ_2, but of θ_1 alone; and λ_{22} must be a function of θ_2 alone.

4.2 The Class of Distributions Admitting Two Unbiassed Statistics of Minimum Variance:- Let us regard

$$\frac{\partial \log \Phi}{\partial \theta_1} = \frac{T_1 - \theta_1}{\lambda_{11}} \quad ; \quad \frac{\partial \log \Phi}{\partial \theta_2} = \frac{T_2 - \theta_2}{\lambda_{22}}$$

as two simultaneous partial differential equations in a function Φ, all the other quantities being given.

Write $\int_0^{\theta_1} \lambda_{11}^{-1}\, d\theta_1 = F_1(\theta_1)$; $\int_0^{\theta_1} F_1(\theta_1)\, d\theta_1 = F_3(\theta_1)$

$\int_0^{\theta_2} \lambda_{22}^{-1}\, d\theta_2 = F_2(\theta_2)$; $\int_0^{\theta_2} F_2(\theta_2)\, d\theta_2 = F_4(\theta_2)$

The functions F_1, F_3 involve θ_1 only ; F_2, F_4 involve θ_2 only, by Section 4.1.2.

Integrating the first partial differential equation, we obtain $\log \Phi = T_1 F_1(\theta_1) + \int \theta_1 \frac{dF_1}{d\theta_1} \cdot d\theta_1$

$$= (T_1 - \theta_1) F_1(\theta_1) + F_3(\theta_1) + c_1'(x' \cdot \theta_2)$$

where c_1' is a constant of integration, independent of θ_1, but possibly involving θ_2 and x'.

Similarly, the second partial differential equation gives $\log \Phi = (T_2 - \theta_2) F_2(\theta_2) + F_4(\theta_2) + c_2'(x' \cdot \theta_1)$

where c_2' is independent of θ_2.

These two expressions for $\log \Phi$ are the same if and only if

$$\left.\begin{aligned} c_1'(x' \cdot \theta_2) &= (T_2 - \theta_2) F_2(\theta_2) + F_4(\theta_2) + c'(x') \\ c_2'(x' \cdot \theta_1) &= (T_1 - \theta_1) F_1(\theta_1) + F_3(\theta_1) + c'(x') \end{aligned}\right\}$$

where c' denotes an arbitrary function of x'. Hence $\log \Phi = (T_1 - \theta_1) F_1(\theta_1) + (T_2 - \theta_2) F_2(\theta_2) + F_3(\theta_1) + F_4(\theta_2) + c'(x')$.

Φ is a product of n similar factors φ. i.e.

$$\Phi(x' \mid \theta_1 \cdot \theta_2) = \prod_{i=1}^{n} \varphi(x_i \mid \theta_1 \cdot \theta_2)$$

We can consequently determine functions $f_1(x)$, $f_2(x)$, $c(x)$ such that $T_1 = \sum_{i=1}^{n} f_1(x_i)$; $T_2 = \sum_{i=1}^{n} f_2(x_i)$; $c'(x') = \sum_{i=1}^{n} c(x_i)$.

Writing

$$U(\theta_1) = \frac{1}{n}\left\{F_3(\theta_1) - \theta_1 F_1(\theta_1)\right\}$$

$$V(\theta_2) = \frac{1}{n}\left\{F_4(\theta_2) - \theta_2 F_2(\theta_2)\right\}$$

we have

$$\log \varphi = F_1(\theta_1) f_1(x) + F_2(\theta_2) f_2(x) + U(\theta_1) + V(\theta_2) + c(x)$$

or

$$\varphi = \exp\left\{F_1(\theta_1) f_1(x) + F_2(\theta_2) f_2(x) + U(\theta_1) + V(\theta_2) + c(x)\right\}$$

The functions F_1, F_2, U, V are known in terms of the original $\lambda_{11}, \lambda_{22}$; and f_1, f_2 are known in terms of the T_1, T_2. The condition that φ should represent a probability distribution over some given range (a, b) (independent of θ_1 and θ_2) is that a real function $c(x)$ should exist which makes

$$\int_a^b \varphi \, dx = 1 .$$

for all values of θ_1 and θ_2. If this is so, φ is a particular case of Koopman's two parameter distribution. In other words, if two unbiassed statistics of minimum variance can exist for a given distribution, the distribution admits a pair of sufficient statistics. By application of Koopman's Second Theorem we deduce, furthermore, that

$$\left.\begin{aligned} T_1 &= \sum_{i=1}^{n} f_1(x_i) \\ T_2 &= \sum_{i=1}^{n} f_2(x_i) \end{aligned}\right\}$$

would themselves be sufficient.

4.3 The General Koopman Distribution of Two Parameters and The Unbiassed Statistics of Minimum Variance :-

Under what conditions can we obtain unbiassed statistics of minimum variance f for the coefficients in

$$\varphi = \exp\{\psi_1 f_1(x) + \psi_2 f_2(x) + F(\psi_1 \cdot \psi_2) + c(x)\} \quad ?$$

[We assume that $\int \varphi \, dx = 1$ for all values of ψ_1, ψ_2.]

In a Bernouillian sample of n observations,

$$\log \Phi = \psi_1 \sum_i f_1(x_i) + \psi_2 \sum_i f_2(x_i) + nF(\psi_1 \cdot \psi_2) + \sum_i c(x_i).$$

Let θ_1, θ_2 be functions of ψ_1 and ψ_2 for which unbiassed statistics of minimum variance can be found. Since

$$\frac{\partial \log \Phi}{\partial \theta_i} = n\left[\frac{1}{n}\sum_j f_1(x_j)\frac{\partial \psi_1}{\partial \theta_i} + \frac{1}{n}\sum_j f_2(x_j)\frac{\partial \psi_2}{\partial \theta_i} + \frac{\partial F}{\partial \theta_i}\right]$$

$[i = 1, 2]$ must be of the form

$$\left(T_i(x') - \theta_i\right) / \lambda_{ii}$$

it is necessary that

$$\frac{\partial \psi_2}{\partial \theta_1} = 0 \quad ; \quad \frac{\partial \psi_1}{\partial \theta_2} = 0.$$

[Alternatively, it is necessary that $\partial \psi_1 / \partial \theta_1 = \partial \psi_2 / \partial \theta_2 = 0$, but this is merely equivalent to pointing out that f_2 might have been labelled f_1, and vice versa]

Now differentiate θ_1 and θ_2, which are functions of ψ_1 and ψ_2, partially with respect to θ_1.

We find

$$1 = \frac{\partial \theta_1}{\partial \psi_1} \cdot \frac{\partial \psi_1}{\partial \theta_1} + \frac{\partial \theta_1}{\partial \psi_2} \cdot \frac{\partial \psi_2}{\partial \theta_1}$$

$$0 = \frac{\partial \theta_2}{\partial \psi_1} \frac{\partial \psi_1}{\partial \theta_1} + \frac{\partial \theta_2}{\partial \psi_2} \frac{\partial \psi_2}{\partial \theta_1}$$

Consequently $\dfrac{\partial \psi_2}{\partial \theta_1} = 0$ implies $\dfrac{\partial \theta_2}{\partial \psi_1} = 0$

Similarly $\frac{\partial \psi_1}{\partial \theta_2} = 0$ implies $\frac{\partial \theta_1}{\partial \psi_2} = 0$.

That is, it is necessary that θ_1, (or θ_2) be a function of ψ_1 only, (or ψ_2 only) say

$$\theta_1 = u(\psi_1) \; ; \quad \theta_2 = v(\psi_2)$$

Then

$$\left(\frac{1}{n} \frac{d\theta_i}{d\psi_i}\right) \frac{\partial \log \Phi}{\partial \theta_i} = \frac{1}{n} \sum_j f_i(x_j) + \frac{\partial F}{\partial \psi_i} \quad (i = 1, 2)$$

To cast this into the desired form of (5) and (6), we choose

$$\frac{\partial F}{\partial \psi_i} = -\theta_i$$

i.e.,

$$\frac{\partial F}{\partial \psi_1} = -u(\psi_1) \; ; \; \frac{\partial F}{\partial \psi_2} = -v(\psi_2).$$

So

$$F = -\int_0^{\psi_1} u(\psi_1)\, d\psi_1 - \int_0^{\psi_2} v(\psi_2)\, d\psi_2$$

$$= U(\psi_1) + V(\psi_2), \text{ say}$$

This is the necessary condition that the Koopman two-parameter distribution admit two unbiassed statistics of minimum variance. The statistics in question are $\frac{1}{n} \sum_{j=1}^{n} f_1(x_j)$ and $\frac{1}{n} \sum_{j=1}^{n} f_2(x_j)$, which estimate $-\partial F/\partial \psi_1 = -dU/d\psi_1$ and $-dV/d\psi_2$ respectively, without bias. The variances by equations 8, 10 are $\frac{1}{n} \frac{d\theta_i}{d\psi_i}$ $(i = 1, 2)$, that is $-\frac{1}{n} \frac{d^2 U}{d\psi_1^2}$ and $-\frac{1}{n} \frac{d^2 V}{d\psi_2^2}$ respectively. The former quantity is a function of ψ_1, (or $\theta_1 = -dU/d\psi_1$) only; the latter of ψ_2 alone - as is required by Section 4.1.2.

4.3.1. <u>Basic Parametric Coefficients</u> :- From the preceding paragraph it is clear that unbiassed statistics of minimum variance can be found only for the coefficients $-d\mathcal{U}(\psi_1)/d\psi_1$ and $-dV(\psi_2)/d\psi_2$, or for linear combinations thereof. These may therefore be termed the basic parametric coefficients for the distribution.

4.3.2. It may be interpolated here that the actual existence of unbiassed statistics of minimum variance for a two parameter distribution has not been proved. All that we have proved is that if there is a distribution

$$\varphi = \exp\left\{\psi_1 f_1(x) + \psi_2 f_2(x) + \mathcal{U}(\psi_1) + V(\psi_2) + c(x)\right\} \tag{11}$$

then it admits two such statistics. Whether such a distribution is compatible with the total probability condition has not yet been investigated. We shall revert to this subject in the next chapter.

4.4. Comparison With Maximum Likelihood:- Assuming the existence of probability distributions of the form (11) their two coefficients can be estimated by Maximum Likelihood. One readily finds that

(i) this methos leads to $\frac{1}{n}\sum_{i=1}^{n} f_1(x_i)$ and $\frac{1}{n}\sum_{i=1}^{n} f_2(x_i)$ as estimates of $-d\mathcal{U}(\psi_1)/d\psi_1$ and of $-dV(\psi_2)/d\psi_2$ respectively. (These particular coefficients being chosen arbitrarily, there being no apparent reason, apart from algebraic simplicity, for preferring them to ψ_1 and ψ_2, or any other functions of ψ_1, ψ_2)

(ii) in indefinitely large samples of size n, the variances of these two statistics are $-\frac{1}{n}\,d^2\mathcal{U}(\psi_1)/d\psi_1^2$

and $-\frac{1}{n} d^2 V(\psi_2)/d\psi_2^2$ respectively: and their covariance is zero. As we know, these values are also valid in finite samples.

In the next chapter, we shall derive some general results, of which the foregoing are merely special cases.

CHAPTER FIVE

THE ESTIMATION OF TWO PARAMETERS BY UNBIASSED STATISTICS OF MINIMUM GENERALISED VARIANCE

5.0 We found in the last chapter that a distribution which admits two unbiassed statistics of minimum variance - if it exists at all - must be of a very special form; so special, it is disconcerting to realise, that it covers none of the common frequency distributions. Why, then, are the criteria, which proved eminently reasonable for estimating one parameter, so exclusive when two coefficients are involved? Partly, at least, the reason is that statistics which satisfy them are uncorrelated. This severe restriction was imposed in the course of minimising the variance of each statistic. A possible way of modifying our standards may therefore lie in seeking to minimise not several, but a single, quantity. For consistency with our preceding work, we should choose a quantity which is in some sense an extension of the notion of variance, and which reduces to the variance itself when only one parameter has to be estimated. Experience in other branches of statistics suggests that the "generalised variance" might possess the requisite properties.

Generalised variance - the concept is due to Frisch and to Wilks - is defined thus: let $T_1, T_2, \cdots\cdots T_r$ be estimates of coefficients $\theta_1, \theta_2, \cdots\cdots \theta_r$ respectively, with variances

$V_{11}, V_{22}, \ldots\ldots V_{\nu\nu}$ respectively. Denote the covariance of T_i and T_j by V_{ij} ($= V_{ji}$ of course). Then the generalised variance of the statistics $\{T\}$ is the determinant of order ν

$$V = |V_{ij}| .$$

Armed with this definition, we tentatively formulate our new requirements for the estimation of several parameters: we want statistics which

(i) are unbiassed

(ii) make the generalised variance a minimum.

We proceed to the application of these criteria, and note that, when only one parameter has to be estimated, they are precisely those studied in Chapter Three.

5.1 Unbiassed Statistics of Minimum Generalised Variance:- Specify the population, from which a sample is to be drawn, by $\varphi(x|\theta_1 \cdot \theta_2)$, a probability function continuous in all three variables, possessing continuous partial derivatives of all orders. The range of the variate x is independent of θ_1, θ_2 though it may be finite or infinite.

We wish to determine statistics T_1, T_2 which

(i) are unbiassed estimates of θ_1 and θ_2 respectively, i.e.,

$$\int (T_1 - \theta_1)\Phi\, dx' = 0 \quad ; \quad \int (T_2 - \theta_2)\Phi\, dx' = 0 .$$

(ii) minimise the generalised variance i.e make

$$V = \begin{vmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{vmatrix} = \text{minimum}$$

where $V_{ii} = \int (T_i - \theta_i)^2 \Phi \, dx' \qquad [i = 1, 2]$

$$V_{12} = V_{21} = \int (T_1 - \theta_1)(T_2 - \theta_2) \Phi \, dx'.$$

In these equations

$$\Phi = \prod_{i=1}^{n} \varphi(x_i \mid \theta_1 \cdot \theta_2) \qquad (1)$$

n being the number of observations $x_i \; [i = 1, 2, \cdots n]$ comprising the sample.

By the total probability condition

$\int \varphi \, dx = 1$ for all valuesoof θ_1 and θ_2

and $\int \Phi \, dx' = 1$ for all values of θ_1 and θ_2.

As in Section 4.0, an unbiassed statistic, for a distribution subject to the multiplicative law (1), is of the form

$$T_i(x') = \theta_i + \frac{1}{\Phi}\left(\lambda_{i1}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{i2}\frac{\partial \Phi}{\partial \theta_2}\right)$$
$$+ \sum_{j=1}^{n}\left\{\frac{1}{\varphi(x_j \mid \theta_1 \cdot \theta_2)}\left(\mu_{i(1\cdot1)}\frac{\partial^2}{\partial \theta_1^2} + 2\mu_{i(1\cdot2)}\frac{\partial^2}{\partial \theta_1 \partial \theta_2} + \mu_{i(2\cdot2)}\frac{\partial^2}{\partial \theta_2^2}\right)\varphi(x_j \mid \theta_1 \cdot \theta_2)\right\}$$
$$+ \sum_{j=1}^{n}\left\{\frac{1}{\varphi(x_j \mid \theta_1 \cdot \theta_2)}\left(\mu_{i(1\cdot1\cdot1)}\frac{\partial^3}{\partial \theta_1^3} + 3\mu_{i(1\cdot1\cdot2)}\frac{\partial^3}{\partial \theta_1^2 \partial \theta_2} + 3\mu_{i(2\cdot2\cdot1)}\frac{\partial^3}{\partial \theta_1 \partial \theta_2^2} + \mu_{i(2\cdot2\cdot2)}\frac{\partial^3}{\partial \theta_2^3}\right)\varphi(x_j \mid \theta_1 \cdot \theta_2)\right\} + \cdots\cdots \qquad (2)$$

$[i = 1, 2]$ The λ's and μ's are functions of θ_1 and θ_2 only, and T_i depends on x' alone.

Let T_i be changed by a small arbitrary amount to $T_i + \epsilon_i f_i(x')$. $[i = 1, 2]$ This changes the generalised variance from V to

$$V' = \begin{vmatrix} V_{11}' & V_{12}' \\ V_{21}' & V_{22}' \end{vmatrix}$$

where

$$V_{11}' = \int \{T_1(x') + \epsilon_1 f_1(x') - \theta_1\}^2 \Phi(x' \mid \theta_1 \cdot \theta_2) \, dx'$$

and so on.

Consequently,

$$V'-V = \left[\int\{T_1(x')+\epsilon_1 f_1(x')-\theta_1\}^2\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]$$
$$\times\left[\int\{T_2(y')+\epsilon_2 f_2(y')-\theta_2\}^2\,\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$
$$-\left[\int\{T_1(x')+\epsilon_1 f_1(x')-\theta_1\}\{T_2(x')+\epsilon_2 f_2(x')-\theta_2\}\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]^2$$
$$-\left[\int\{T_1(x')-\theta_1\}^2\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]\left[\int\{T_2(y')-\theta_2\}^2\,\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$
$$+\left[\int\{T_1(x')-\theta_1\}\{T_2(x')-\theta_2\}\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]^2 \qquad (3)$$

When ϵ_1, ϵ_2 are so small that terms in ϵ_1^2, ϵ_2^2, $\epsilon_1\epsilon_2$ and in higher powers are negligible, (3) reduces to

$$\delta V = 2\epsilon_1\left[\int\{T_1(x')-\theta_1\}f_1(x')\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]$$
$$\times\left[\int\{T_2(y')-\theta_2\}^2\,\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$
$$+2\epsilon_2\left[\int\{T_2(y')-\theta_2\}f_2(y')\,\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$
$$\times\left[\int\{T_1(x')-\theta_1\}^2\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]$$
$$-\epsilon_1\left[\int\{T_2(x')-\theta_2\}f_1(x')\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]$$
$$\times\left[\int\{T_1(y')-\theta_1\}\{T_2(y')-\theta_2\}\,\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$
$$-\epsilon_2\left[\int\{T_1(x')-\theta_1\}f_2(x')\,\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]$$
$$\times\left[\int\{T_1(y')-\theta_1\}\{T_2(y')-\theta_2\}\,\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$

$$-\epsilon_1\left[\int\{T_2(y')-\theta_2\}f_1(y')\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$
$$\times\left[\int\{T_1(x')-\theta_1\}\{T_2(x')-\theta_2\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]$$
$$-\epsilon_2\left[\int\{T_1(y')-\theta_1\}f_2(y')\Phi(y'|\theta_1\cdot\theta_2)\,dy'\right]$$
$$\times\left[\int\{T_1(x')-\theta_1\}\{T_2(x')-\theta_2\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'\right]$$

or, remembering the definition of V_{11}, V_{22}, V_{12},

$$\delta V = 2\epsilon_1\int\Phi(x'|\theta_1\cdot\theta_2)f_1(x')\,dx'$$
$$\times\left[\{T_1(x')-\theta_1\}V_{22}-\{T_2(x')-\theta_2\}V_{12}\right]$$
$$+2\epsilon_2\int\Phi(x'|\theta_1\cdot\theta_2)f_2(x')\,dx'$$
$$\times\left[\{T_2(x')-\theta_2\}V_{11}-\{T_1(x')-\theta_1\}V_{21}\right]$$

We restrict the arbitrary variations f_1, f_2 as in Section 4.0, so that

$$f_1(x') = \sum_{r_1} u_{r_1} + \sum_s b_s u_s$$

$$f_2(x') = \sum_{r_2} u_{r_2} + \sum_s c_s v_s .$$

Here (i)
$$u_{r_i} = \sum_{j=0}^{r_i}\frac{a_{r_i-j\cdot j}}{\Phi}\cdot\frac{\partial^{r_i}\Phi}{\partial\theta_1^{\,r_i-j}\,\partial\theta_2^{\,j}} \qquad [i=1,2]$$

and is evaluated for the actual population values of θ_1 and θ_2. The a's are arbitrary constants

(ii) u_s is any solution of

$$\int u(x') \Phi(x' | \theta_1 \cdot \theta_2)\, dx' = h_s(\theta_2)$$

and v_s any solution of

$$\int v(x') \Phi(x' | \theta_1 \cdot \theta_2)\, dx' = h'_s(\theta_1) \quad ,$$

h_s, h'_s denoting any continuous functions of θ_2 and θ_1 respectively.

(iii) the b's and c's are arbitrary numerical constants, and the various summations cover as many terms as we please.

When T_1, T_2 are given by (2), it is readily seen that a necessary condition for stationary generalised variance, or for $\delta V = 0$ is

$$\mu_{i(1\cdot1)} = \mu_{i(1\cdot2)} = \mu_{i(2\cdot2)} = \mu_{i(1\cdot1\cdot1)} = \cdots\cdots = 0 \qquad [i = 1, 2]$$

Accordingly, we need only consider whether unbiassed statistics of the form

$$T_i \equiv \theta_i + \frac{1}{\Phi}\left(\lambda_{i1} \frac{\partial \Phi}{\partial \theta_1} + \lambda_{i2} \frac{\partial \Phi}{\partial \theta_2}\right) \cdot [i = 1, 2] \quad (4)$$

can make $\delta V = 0$. Anticipating somewhat, we can show that when the statistics are given by (4),

$$V_{11} = \lambda_{11} \;;\quad V_{22} = \lambda_{22} \;;\quad V_{12} = V_{21} = \lambda_{12} = \lambda_{21} \quad . \quad (5)$$

Substitution of (4) and (5) into the expression for δV yields

$$\delta V = 2\epsilon_1 \int f_1(x')\, dx' \left[\left(\lambda_{11}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{12}\frac{\partial \Phi}{\partial \theta_2}\right)\lambda_{22} - \left(\lambda_{21}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{22}\frac{\partial \Phi}{\partial \theta_2}\right)\lambda_{12}\right]$$

$$+ 2\epsilon_2 \int f_2(x')\, dx' \left[\left(\lambda_{21}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{22}\frac{\partial \Phi}{\partial \theta_2}\right)\lambda_{11} - \left(\lambda_{11}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{12}\frac{\partial \Phi}{\partial \theta_2}\right)\lambda_{21}\right]$$

$$= \left(\lambda_{11}\lambda_{22} - \lambda_{12}^2\right)\left[2\epsilon_1 \frac{\partial}{\partial \theta_1}\int f_1(x')\Phi\, dx' + 2\epsilon_2 \frac{\partial}{\partial \theta_2}\int f_2(x')\Phi\, dx'\right]$$

Now, from the definition of the f's

$$\int f_1(x')\Phi\,dx' = \sum_s b_s \int u_s(x')\Phi\,dx' = \sum_s b_s h_s(\theta_2)$$

since $\int \frac{\partial^r \Phi}{\partial\theta_1^{r-j}\,\partial\theta_2^{j}}\,dx' = 0$

Consequently, $\frac{\partial}{\partial\theta_1}\int f_1(x')\Phi\,dx' = 0$

Similarly $\frac{\partial}{\partial\theta_2}\int f_2(x')\Phi\,dx' = 0$

Therefore $\delta V = 0$ and a pair of unbiassed statistics of the form (4), makes the generalised variance stationary.

5.1.1 Nature of the Stationary Value of the Generalised Variance:- Intuitively, one might expect that the stationary value of the generalised variance was a minimum. This will be so if our statistics make the second variation of V always positive. Since this second variation is the set of quadratic terms in (ϵ_1, ϵ_2) in (3), we have

$$\begin{aligned}
\delta_2 V = {} & \epsilon_1^2\int f_1^2(x')\Phi(x'|\theta_1\cdot\theta_2)\,dx'\int\{T_2(y')-\theta_2\}^2\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& + \epsilon_2^2\int f_2^2(y')\Phi(y'|\theta_1\cdot\theta_2)\,dy'\int\{T_1(x')-\theta_1\}^2\Phi(x'|\theta_1\cdot\theta_2)\,dx' \\
& + 4\epsilon_1\epsilon_2\int f_1(x')\{T_1(x')-\theta_1\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'\int f_2(y')\{T_2(y')-\theta_2\}\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& - \epsilon_1^2\int f_1(x')\{T_2(x')-\theta_2\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'\int f_1(y')\{T_2(y')-\theta_2\}\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& - \epsilon_2^2\int f_2(x')\{T_1(x')-\theta_1\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'\int f_2(y')\{T_1(y')-\theta_1\}\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& - \epsilon_1\epsilon_2\int f_1(x')f_2(x')\Phi(x'|\theta_1\cdot\theta_2)\,dx'\int\{T_1(y')-\theta_1\}\{T_2(y')-\theta_2\}\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& - \epsilon_1\epsilon_2\int f_1(x')\{T_2(x')-\theta_2\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'\int f_2(y')\{T_1(y')-\theta_1\}\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& - \epsilon_1\epsilon_2\int f_2(x')\{T_1(x')-\theta_1\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'\int f_1(y')\{T_2(y')-\theta_2\}\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& - \epsilon_1\epsilon_2\int f_1(y')f_2(y')\Phi(y'|\theta_1\cdot\theta_2)\,dy'\int\{T_1(x')-\theta_1\}\{T_2(x')-\theta_2\}\Phi(x'|\theta_1\cdot\theta_2)\,dx'
\end{aligned}$$

$$= \epsilon_1^2 V_{22} \int f_1^2(x') \Phi(x'|\theta_1 \cdot \theta_2)\, dx' + \epsilon_2^2 V_{11} \int f_2^2(x') \Phi(x'|\theta_1 \cdot \theta_2)\, dx'$$

$$+ 4\epsilon_1\epsilon_2 \int f_1(x')\{T_1(x') - \theta_1\}\Phi(x'|\theta_1 \cdot \theta_2)\, dx' \int f_2(y')\{T_2(y') - \theta_2\}\Phi(y'|\theta_1 \cdot \theta_2)\, dy'$$

$$- \epsilon_1^2\left[\int f_1(x')\{T_2(x') - \theta_2\}\Phi(x'|\theta_1 \cdot \theta_2)\, dx'\right]^2 - \epsilon_2^2\left[\int f_2(x')\{T_1(x') - \theta_1\}\Phi(x'|\theta_1 \cdot \theta_2)\, dx'\right]^2$$

$$- 2\epsilon_1\epsilon_2 V_{12} \int f_1(x') f_2(x') \Phi(x'|\theta_1 \cdot \theta_2)\, dx'$$

$$- 2\epsilon_1\epsilon_2 \int f_1(x')\{T_2(x') - \theta_2\}\Phi(x'|\theta_1 \cdot \theta_2)\, dx' \int f_2(y')\{T_1(y') - \theta_1\}\Phi(y'|\theta_1 \cdot \theta_2)\, dy'$$

Inserting the values of T_1, T_2 as given by (4), and remembering (5) and the relations

$$\int f_1(x') \frac{\partial \Phi}{\partial \theta_1}\, dx' = \int f_2(x') \frac{\partial \Phi}{\partial \theta_2}\, dx' = 0$$

we obtain

$$\delta_2 V = \epsilon_1^2\left[\lambda_{22}\int f_1^2 \Phi\, dx' - \lambda_{22}^2\left(\int f_1 \frac{\partial \Phi}{\partial \theta_2}\, dx'\right)^2\right]$$

$$+ \epsilon_2^2\left[\lambda_{11}\int f_2^2 \Phi\, dx' - \lambda_{11}^2\left(\int f_2 \frac{\partial \Phi}{\partial \theta_1}\, dx'\right)^2\right]$$

$$- 2\epsilon_1\epsilon_2\left[\lambda_{12}\int f_1 f_2 \Phi\, dx' + (\lambda_{11}\lambda_{22} - 2\lambda_{12}^2)\int f_1 \frac{\partial \Phi}{\partial \theta_2}\, dx' \int f_2 \frac{\partial \Phi}{\partial \theta_1}\, dy'\right]$$

In special cases, $\delta_2 V$ is > 0 (e.g., when $f_2 = 0$ and $f_1 = \sum \mu_{r_1}$, the latter implying $\int f_1 (\partial\Phi/\partial\theta_2)\, dx' = 0$) It is consequently always positive if the discriminant of the foregoing quadratic form is negative, i.e., if

$$\left[\lambda_{22}\int f_1^2 \Phi\, dx' - \lambda_{22}^2\left(\int f_1 \frac{\partial \Phi}{\partial \theta_2}\, dx'\right)^2\right]\left[\lambda_{11}\int f_2^2 \Phi\, dx' - \lambda_{11}^2\left(\int f_2 \frac{\partial \Phi}{\partial \theta_1}\, dx'\right)^2\right]$$

$$> \left[\lambda_{12}\int f_1 f_2 \Phi\, dx' + (\lambda_{11}\lambda_{22} - 2\lambda_{12}^2)\int f_1 \frac{\partial \Phi}{\partial \theta_2}\, dx' \int f_2 \frac{\partial \Phi}{\partial \theta_1}\, dy'\right]^2 \qquad (6)$$

Assume that the generalised variance of T_1, T_2 is positive, i.e., (equation (5)) that $\lambda_{11}\lambda_{22} > \lambda_{12}^2$.

<u>The Second Variation in Large Samples</u>:- A closer inequality than (6) may be derived by means of the following

Lemma:- For any real, continuous, functions

$$\begin{vmatrix} \int f_1^2(x')\Phi(x')\,dx' & \int f_1(x')f_2(x')\Phi(x')\,dx' \\ \int f_1(x')f_2(x')\Phi(x')\,dx' & \int f_2^2(x')\Phi(x')\,dx' \end{vmatrix}$$

$$= \frac{1}{2}\iint \left\{ f_1(x')f_2(y') - f_1(y')f_2(x') \right\}^2 \Phi(x')\Phi(y')\,dx'\cdot dy'$$

Obviously, when we have a finite number of terms r,

$$\begin{vmatrix} \sum_{i=1}^{r} h f_{1i}^2 \Phi_i & \sum_{i=1}^{r} h f_{1i} f_{2i} \Phi_i \\ \sum_{i=1}^{r} h f_{1i} f_{2i} \Phi_i & \sum_{i=1}^{r} h f_{2i}^2 \Phi_i \end{vmatrix}$$

$$= \frac{1}{2} \sum_{i,j=1}^{r}{}' h^2 \left\{ f_{1i} f_{2j} - f_{1j} f_{2i} \right\}^2 \Phi_i \Phi_j \, .$$

where in $\sum'$ the terms $i=j$ are omitted. But, when $i=j$ the term $\{f_{1i} f_{2j} - f_{1j} f_{2i}\}$ is zero, and $\sum'$ can be replaced by $\sum_{i=1}^{r}\sum_{j=1}^{r}$. Proceeding to the limit, $r \to \infty$, $h \to 0$ the sums tend to integrals, and the lemma is established.

Returning to (6), write

$$E \equiv \lambda_{11}\lambda_{22}\int f_1^2\Phi\,dx'\int f_2^2\Phi\,dx' - \lambda_{12}^2\left(\int f_1 f_2 \Phi\,dx'\right)^2 .$$

Since

$$\lambda_{11}\lambda_{22} > \lambda_{12}^2 \quad ; \quad (\lambda_{11} > 0 \, , \ \lambda_{22} > 0) \, ,$$

$$E > \lambda_{11}\lambda_{22} \begin{vmatrix} \int f_1^2\Phi\,dx' & \int f_1 f_2 \Phi\,dx' \\ \int f_1 f_2 \Phi\,dx' & \int f_2^2 \Phi\,dx' \end{vmatrix}$$

$$> \frac{1}{2}\lambda_{11}\lambda_{22}\iint\left\{ f_1(x')f_2(y') - f_1(y')f_2(x')\right\}^2 \Phi(x'|\theta_1\theta_2)\Phi(y'|\theta_1\theta_2)\,dx'dy'$$

$\therefore \quad E > 0$

The inequality (6) can be rearranged thus:

$$
\begin{aligned}
I_s E > {} & \lambda_{11}\lambda_{22}^2\left(\int f_1 \frac{\partial \Phi}{\partial \theta_2}\,dx'\right)^2 \int f_2^2 \Phi\,dy' \\
& + \lambda_{11}^2\lambda_{22}\left(\int f_2 \frac{\partial \Phi}{\partial \theta_1}\,dx'\right)^2 \int f_1^2 \Phi\,dy' \\
& - 4\lambda_{12}^2\left(\lambda_{11}\lambda_{22}-\lambda_{12}^2\right)\left(\int f_1 \frac{\partial \Phi}{\partial \theta_2}\,dx'\right)^2\left(\int f_2 \frac{\partial \Phi}{\partial \theta_1}\,dy'\right)^2 \\
& + 2\lambda_{12}\left(\lambda_{11}\lambda_{22}-2\lambda_{12}^2\right)\int f_1 f_2 \Phi\,dx' \int f_1 \frac{\partial \Phi}{\partial \theta_2}\,dy' \int f_2 \frac{\partial \Phi}{\partial \theta_1}\,dz' \; ?
\end{aligned} \tag{6'}
$$

The third term on the right hand side is essentially negative. Consequently (6) is certainly true if

$$
\begin{aligned}
& \frac{1}{2}\iint\left\{f_1(x')f_2(y') - f_2(x')f_1(y')\right\}^2 \Phi(x'|\theta_1\cdot\theta_2)\,\Phi(y'|\theta_1\cdot\theta_2)\,dx'\,dy' \\
& > \lambda_{22}\left(\int f_1(x')\frac{\partial \Phi}{\partial \theta_2}\,dx'\right)^2\cdot\int f_2^2(y')\,\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& + \lambda_{11}\left(\int f_2(x')\frac{\partial \Phi}{\partial \theta_1}\,dx'\right)^2\cdot\int f_1^2(y')\,\Phi(y'|\theta_1\cdot\theta_2)\,dy' \\
& +
\end{aligned}
$$

$$+ \left[2\left\{ \left| \lambda_{12} \left(\lambda_{11}\lambda_{22} - 2\lambda_{12}^2 \right) \right| \right\} / \lambda_{11}\lambda_{22} \right.$$
$$\left. \times \int f_1(x') f_2(x') \Phi(x' | \theta_1 \cdot \theta_2)\, dx' \int f_1(y') \frac{\partial \Phi}{\partial \theta_2}\, dy' \int f_2(z') \frac{\partial \Phi}{\partial \theta_1}\, dz' \right]$$

\- the notation $|\ |$ denotes absolute value. Now this inequality is true in sufficiently large samples. By their definition f_1, f_2 do not necessarily tend to zero as the size of the sample n increases. Therefore, when f_1, f_2 have been specified in any particular case, the left hand side of the above inequality has a definite, positive value. On the other hand, as we demonstrate in Section 5.3.2,

$$\lambda_{11} = \lambda_{22} = \lambda_{12} = O(n^{-1})$$

as the sample becomes indefinitely large. By choosing n great enough, we can thus make the right hand side of our inequality smaller than any preassigned quantity. It follows that, as $n \to \infty$, (6) is always valid; that is, $\delta_2 V$ is always > 0 without any additional restriction on the functions f_1, f_2. In the language of the Calculus of Variations, the stationary value of the generalised variance is a strong minimum in indefinitely large samples.

The Second Variation in Finite Samples:- An example will illustrate that the inequality (6) is not universally true.

Choose $f_1(x') = \sum u_r$ so that

$\int f_1(\partial\Phi/\partial\theta_2)\, dx' = 0$ and (6) reduces to

Is $\lambda_{22}\left[\int f_1^2 \Phi\, dx'\right]\left[\lambda_{11}\int f_2^2 \Phi\, dx' - \lambda_{11}^2\left(\int f_2 \frac{\partial\Phi}{\partial\theta_1} dx'\right)^2\right] > \lambda_{12}^2\left(\int f_1 f_2 \Phi\, dx'\right)^2$?

Make now two further restrictions, viz.,

(i) $f_1(x') = \sum_r \frac{a_{0\cdot r}}{\Phi} \cdot \frac{\partial^r \Phi}{\partial \theta_2^r}$ evaluated for the population values of θ_1, θ_2 (in the general expression for μ_r we have thus put $a_{r-j\cdot j} = 0$ when $j \neq r$)

and (ii) $f_2 = w$ a solution of

$$\int w(x') \Phi(x' | \theta_1 \cdot \theta_2)\, dx' = h'(\theta_1)$$

where h' is a function of θ_1, to be more closely defined later. These restrictions give

$$\int f_1 f_2 \Phi\, dx' = \int (f_1 \Phi) f_2\, dx' = \sum_r a_{0\cdot r} \int w \frac{\partial^r \Phi}{\partial \theta_2^r}\, dx'$$

$$= \sum_r a_{0\cdot r} \frac{\partial^r}{\partial \theta_2^r} \int w \Phi\, dx' = \sum_r a_{0\cdot r} \frac{\partial^r h'(\theta_1)}{\partial \theta_2^r} = 0$$

and the foregoing inequality becomes

Is $\lambda_{22} \int f_1^2 \Phi\, dx' \left[\lambda_{11} \int f_2^2 \Phi\, dx' - \lambda_{11}^2 \left(\int f_2 (\partial \Phi / \partial \theta_1)\, dx' \right)^2 \right] > 0$?

or, since the variances λ_{11} and λ_{22}, and $\int f_1^2 \Phi\, dx'$ are all positive,

Is $\frac{1}{\lambda_{11}} \int w^2 \Phi\, dx' > \left[\int w (\partial \Phi / \partial \theta_1)\, dx' \right]^2$

$> \left\{ dh'(\theta_1) / d\theta_1 \right\}^2$? (7)

To prove that this is not necessarily true, consider the following:<u>Lemma</u>: If $|h'(\theta_1)| < K$ (a constant) for all values of θ_1, then $|w(x')| < K$ for all values of $x' \equiv (x_1 \cdots x_n)$ where $\int w(x') \Phi(x' | \theta_1 \cdot \theta_2)\, dx' = h'(\theta_1)$, and h' is continuous.

Since $\Phi > 0$ everywhere, the Mean Value Theorem for integrals gives

$$h'(\theta_1) = \int w(x') \Phi(x' | \theta_1 \cdot \theta_2)\, dx' = w(\xi') \int \Phi(x' | \theta_1 \cdot \theta_2)\, dx' = w(\xi')$$

ξ' is a vector $(\xi_1, \xi_2 \cdots \xi_n)$ such that $a \leq \xi_i \leq b$ $[i = 1, 2, \cdots n]$ where a, b are the limits of the variate x. The particular value of ξ' depends on the particular value of θ_1, a point which we stress by writing the last equation as

$$h'(\theta_1) = \omega(\xi'; \theta_1)$$

$\omega(\xi'; \theta_1)$ is a continuous function of θ_1, for

$$\lim_{\theta_1^x \to \theta_1} \{\omega(\xi'; \theta_1^x) - \omega(\xi'; \theta_1)\} = \lim_{\theta_1^x \to \theta_1} \left[\int \omega(x') \Phi(x' | \theta_1^x \theta_2) dx' - \int \omega(x') \Phi(x' | \theta_1 \cdot \theta_2) dx'\right]$$

$$= \lim_{\theta_1^x \to \theta_1} \int \omega(x') \{ \Phi(x' | \theta_1^x \cdot \theta_2) - \Phi(x' | \theta_1 \cdot \theta_2) \} dx'$$

$$= \lim_{\theta_1^x \to \theta_1} \int \omega(x') (\theta_1^x - \theta_1) \frac{\partial \Phi\{x' | \theta_1 + \alpha(\theta_1^x - \theta_1) \cdot \theta_2\}}{\partial \theta_1} dx' \quad [0 < \alpha < 1]$$

(since Φ is continuous in θ_1)

$$= \lim_{\theta_1^x \to \theta_1} (\theta_1^x - \theta_1) \frac{\partial}{\partial \theta_1} \int \omega(x') \Phi\{x' | \theta_1 + \alpha(\theta_1^x - \theta_1) \cdot \theta_2\} dx'$$

$$= \frac{dh'(\theta_1)}{d\theta_1} \lim_{\theta_1^x \to \theta_1} (\theta_1^x - \theta_1) = 0 .$$

Consequently $\omega(\xi'; \theta_1)$ assumes <u>every</u> value between $\omega(\xi'; \beta)$ and $\omega(\xi'; \gamma)$ as θ_1 assumes every value between its extremes β, γ (Whittaker and Watson, 4th Edition, p.42); and

$$|\omega(\xi'; \theta_1)| = |h'(\theta_1)| < K$$

for all values of ξ', which proves the lemma.

If $|h'(\theta_1)| < K$ always, (7) implies, by our lemma,

$$\left(\frac{dh'(\theta_1)}{d\theta_1}\right)^2 < \frac{1}{\lambda_{11}} \omega^2(\eta') \int \Phi \, dx' \text{ (by the mean value theorem)}$$

$$< K^2 / \lambda_{11}$$

However, as is well known, continuous functions exist which are arbitrarily small, but whose gradients are as large as we please. That is, the last quoted inequality is contradicted by certain continuous functions $h'(\theta_1)$. Thus, unless $dh'(\theta_1)/d\theta_1$ is small enough, the statistics T_1, T_2 may not have a generalised variance less than that of other permitted statistics $T_1 + \epsilon_1 f_1(x')$, $T_2 + \epsilon_2 f_2(x')$.

$\delta_2 V$ will be > 0 always, provided we restrict $f_1(x'), f_2(x')$ in such a way that (6) or $(6')$ is always valid. If the size of the sample (and hence the λ's) is fixed, $(6')$ will be true so long as the gradients of the arbitrary $h'(\theta_1), h(\theta_2)$ are sufficiently small. We therefore say that the statistics make the generalised variance a weak minimum in finite samples. The analogy with a similar result in the ordinary calculus of variations is noteworthy.

5.1.2 Summary:- Given a distribution $\varphi(x \mid \theta_1 \cdot \theta_2)$ let there exist functions $\lambda_{11}(\theta_1 \cdot \theta_2), \lambda_{12}(\theta_1 \cdot \theta_2), \lambda_{21}(\theta_1 \cdot \theta_2), \lambda_{22}(\theta_1 \cdot \theta_2)$ such that

$$\left.\begin{aligned} T_1 &\equiv \theta_1 + \frac{1}{\Phi}\left(\lambda_{11}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{12}\frac{\partial \Phi}{\partial \theta_2}\right) \\ T_2 &\equiv \theta_2 + \frac{1}{\Phi}\left(\lambda_{21}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{22}\frac{\partial \Phi}{\partial \theta_2}\right) \end{aligned}\right\} \quad (8)$$

are functions of x' alone. Then T_1, T_2 are unbiassed estimates of θ_1, θ_2 respectively which make the generalised variance a minimum. The minimum is strong in indefinitely large samples, but weak in finite samples.

5.1.3 Uniqueness:- Define the "generalised second moment about the origin" as

$$v = \begin{vmatrix} V_{11} + \theta_1^2 & V_{12} + \theta_1\theta_2 \\ V_{12} + \theta_1\theta_2 & V_{22} + \theta_2^2 \end{vmatrix}$$

in which the $(i,j)^{th}$ element is the $(i,j)^{th}$ bivariate second moment about the origin, or $\int T_i T_j \Phi\, dx'$.

It is readily found that, when the statistics (8) exist for a Bernouillian sample, the same functions T_1, T_2 enjoy the properties of unbiassedness and minimum generalised variance in a Poissonian sample, and are then given by

$$T_i \equiv \bar{\theta}_i + \sum_{j=1}^{n} \frac{\lambda_{i1}(\theta_{1j}\cdot\theta_{2j})}{\Phi(x'|\theta_1'\cdot\theta_2')}\cdot\frac{\partial\Phi(x'|\theta_1'\theta_2')}{\partial\theta_{1j}} + \frac{\lambda_{i2}(\theta_{1j}\cdot\theta_{2j})}{\Phi(x'|\theta_1'\cdot\theta_2')}\cdot\frac{\partial\Phi(x'\cdot\theta_1'\cdot\theta_2')}{\partial\theta_{2j}}$$

(θ_{ij} ; $j=1,2,\cdots n$; are the values assumed by θ_i in the course of the n observations ; $i=1,2$)

Let the generalised second moment about the origin of T_1, T_2 in a Poissonian sample be $v(\theta_1'\cdot\theta_2')$. That is

$$\begin{vmatrix} \int T_1^2(x')\,\Phi(x'|\theta_1'\theta_2')\,dx' & \int T_1(x')T_2(x')\,\Phi(x'|\theta_1'\cdot\theta_2')\,dx' \\ \int T_1(x')T_2(x')\,\Phi(x'|\theta_1\cdot\theta_2)\,dx' & \int T_2^2(x')\,\Phi(x'|\theta_1'\cdot\theta_2')\,dx' \end{vmatrix} = v(\theta_1'\cdot\theta_2')$$

or

$$\iint\left\{T_1^2(x')T_2^2(y') - T_1(x')T_1(y')T_2(x')T_2(y')\right\}\Phi(x'|\theta_1'\cdot\theta_2')\Phi(y'|\theta_1'\cdot\theta_2')\,dx'dy'$$
$$= v(\theta_1'\cdot\theta_2')$$

Consider now the $2n$-fold linear integral equation with kernel $\Phi(x'|\theta_1'\cdot\theta_2')\,\Phi(y'|\theta_1'\cdot\theta_2')$

$$\iint \mathcal{U}(x'\cdot y')\,\Phi(x'|\theta_1'\cdot\theta_2')\,\Phi(y'|\theta_1'\cdot\theta_2')\,dx'dy' = v(\theta_1'\cdot\theta_2')$$

which has only one solution of class L^2 for $\mathcal{U}(x'\cdot y')$

We know that the continuous function

$$\mathcal{U}(x'\cdot y') \equiv T_1^2(x')T_2^2(y') - T_1(x')T_1(y')T_2(x')T_2(y')$$

satisfies this equation; therefore it is the only continuous solution. That is, no other statistics give the same generalised second moment about the origin as T_1, T_2 in a Poissonian sample. It follows that no other estimates of θ_1, θ_2 have the same generalised variance as T_1, T_2

5.1.4 A Relation Between The λ's :- Solving (8) as linear equations in $\partial \log \Phi / \partial \theta_1$, $\partial \log \Phi / \partial \theta_2$ yields

$$\frac{\partial \log \Phi}{\partial \theta_1} = \alpha_{22}(T_1 - \theta_1) - \alpha_{12}(T_2 - \theta_2)$$

$$\frac{\partial \log \Phi}{\partial \theta_2} = -\alpha_{21}(T_1 - \theta_1) + \alpha_{11}(T_2 - \theta_2)$$

where $\alpha_{ij} = \lambda_{ij} / (\lambda_{11}\lambda_{22} - \lambda_{12}\lambda_{21})$. $[i, j = 1, 2]$

Differentiating with respect to θ_1 and θ_2

$$\frac{\partial^2 \log \Phi}{\partial \theta_1^2} = -\alpha_{22} + (T_1 - \theta_1)\frac{\partial \alpha_{22}}{\partial \theta_1} - (T_2 - \theta_2)\frac{\partial \alpha_{12}}{\partial \theta_1}$$

$$\frac{\partial^2 \log \Phi}{\partial \theta_2^2} = -(T_1 - \theta_1)\frac{\partial \alpha_{21}}{\partial \theta_2} - \alpha_{11} + (T_2 - \theta_2)\frac{\partial \alpha_{11}}{\partial \theta_2}$$

$$\frac{\partial^2 \log \Phi}{\partial \theta_2 \partial \theta_1} = (T_1 - \theta_1)\frac{\partial \alpha_{22}}{\partial \theta_2} + \alpha_{12} - (T_2 - \theta_2)\frac{\partial \alpha_{12}}{\partial \theta_2}$$

$$\frac{\partial^2 \log \Phi}{\partial \theta_1 \partial \theta_2} = \alpha_{21} - (T_1 - \theta_1)\frac{\partial \alpha_{21}}{\partial \theta_1} + (T_2 - \theta_2)\frac{\partial \alpha_{11}}{\partial \theta_1}$$

Multiply each equation by Φ and integrate with respect to x'. Remembering the unbiassedness of T_1, T_2 we obtain

$$\left.\begin{aligned} \int \Phi \frac{\partial^2 \log \Phi}{\partial \theta_1^2}\, dx' &= -\alpha_{22} \\ \int \Phi \frac{\partial^2 \log \Phi}{\partial \theta_2^2}\, dx' &= -\alpha_{11} \end{aligned}\right\} \qquad (9)$$

$$\left.\begin{aligned}\int \Phi \frac{\partial^2 \log \Phi}{\partial\theta_1 \partial\theta_2}\, dx' &= \alpha_{12} \\ &= \alpha_{21} = \int \Phi \frac{\partial^2 \log \Phi}{\partial\theta_2 \partial\theta_1}\, dx'\end{aligned}\right\} \quad (9)$$

From the last line, it follows that

$$\lambda_{12} = \lambda_{21}$$

5.1.5 Variances and Covariance of the Unbiassed Statistics of Minimum Generalised Variance:- The variance of T_1 is

$$V_{11} = \int (T_1 - \theta_1)^2 \, \Phi \, dx'$$
$$= \int \left\{ \lambda_{11}^2 \left(\frac{\partial\Phi}{\partial\theta_1}\right)^2 + 2\lambda_{11}\lambda_{12} \frac{\partial\Phi}{\partial\theta_1} \cdot \frac{\partial\Phi}{\partial\theta_2} + \lambda_{12}^2 \left(\frac{\partial\Phi}{\partial\theta_2}\right)^2 \right\} \frac{dx'}{\Phi}$$

Now

$$\Phi \frac{\partial^2 \log \Phi}{\partial\theta_i \partial\theta_j} = \frac{\partial^2 \Phi}{\partial\theta_i \partial\theta_j} - \frac{\partial\Phi}{\partial\theta_i} \cdot \frac{\partial\Phi}{\partial\theta_j} \cdot \frac{1}{\Phi}$$

and $\int \frac{\partial^2 \Phi}{\partial\theta_i \partial\theta_j}\, dx' = 0$ for all values of θ_1, θ_2

$[\, i, j = 1, 2 \,]$

Therefore

$$V_{11} = -\int\left\{\lambda_{11}^2 \frac{\partial^2 \log \Phi}{\partial \theta_1^2} + 2\lambda_{11}\lambda_{12}\frac{\partial^2 \log \Phi}{\partial \theta_1 \partial \theta_2} + \lambda_{12}^2 \frac{\partial^2 \log \Phi}{\partial \theta_2^2}\right\}\Phi\, dx'$$

or, from (9),

$$V_{11} = \lambda_{11}^2 \alpha_{22} - 2\lambda_{11}\lambda_{12}\alpha_{12} + \lambda_{12}^2 \alpha_{11}.$$

Hence, since $\alpha_{ij} = \lambda_{ij}/(\lambda_{11}\lambda_{22}-\lambda_{12}\lambda_{21})$ and since $\lambda_{12} = \lambda_{21}$,

$$V_{11} = \lambda_{11} \tag{10}$$

Similarly, the variance of T_2 is

$$V_{22} = \lambda_{22} \tag{11}$$

and the covariance of T_1 and T_2 is

$$V_{12} = V_{21} = \lambda_{12} = \lambda_{21} \tag{12}$$

These last three results have been quoted already (equation 5) .

The generalised variance has the value

$$V = \begin{vmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{vmatrix} = \begin{vmatrix} \lambda_{11} & \lambda_{21} \\ \lambda_{12} & \lambda_{22} \end{vmatrix} \tag{13}$$

<u>5.2</u> <u>Example of The Normal Distribution</u>:- We exemplify the general results by considering the distribution

$$\varphi = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$$

whence $\Phi(x' \mid m, \sigma^2) = (2\pi)^{-\frac{n}{2}}\sigma^{-n} \exp\left[\{-\Sigma x^2 + 2m\Sigma x - nm^2\}/2\sigma^2\right]$

Unbiassed estimates of minimum generalised variance exist for m and σ^2 if we can find functions (of m, σ^2) λ_{11}, $\lambda_{12} = \lambda_{21}$, λ_{22} such that

$\lambda_{11}\frac{\partial \log \Phi}{\partial m} + \lambda_{12}\frac{\partial \log \Phi}{\partial \sigma^2} =$ (a function of x') $- m = T_1 - m$, say

$\lambda_{21}\frac{\partial \log \Phi}{\partial m} + \lambda_{22}\frac{\partial \log \Phi}{\partial \sigma^2} =$ (a function of x') $- \sigma^2 = T_2 - \sigma^2$, say

i.e., such that

$$T_1 - m \equiv \lambda_{11}\left[\frac{\sum x}{\sigma^2} - \frac{mn}{\sigma^2}\right] + \frac{\lambda_{12}}{2\sigma^2}\left[-n + \frac{\sum x^2}{\sigma^2} - \frac{2m\sum x}{\sigma^2} + \frac{m^2 n}{\sigma^2}\right]$$

$$T_2 - \sigma^2 \equiv \lambda_{21}\left[\frac{\sum x}{\sigma^2} - \frac{mn}{\sigma^2}\right] + \frac{\lambda_{22}}{2\sigma^2}\left[-n + \frac{\sum x^2}{\sigma^2} - \frac{2m\sum x}{\sigma^2} + \frac{m^2 n}{\sigma^2}\right]$$

The former is satisfied by $\lambda_{11} = \sigma^2/n$, $\lambda_{12} = 0$. Then $\lambda_{21} = 0$ also, and no suitable λ_{22} exists. It is readily seen that no other values of $\lambda_{11}, \lambda_{12}$ ever lead to the mutual satisfaction of this pair of identities. We conclude that the coefficients m and σ^2 do not admit of estimation by unbiassed statistics of minimum generalised variance.

Let us now try to estimate the alternative coefficients m and $(m^2+\sigma^2)$. On writing $\theta_1 = m$, $\theta_2 = m^2+\sigma^2$ we obtain

$$\log \Phi = \frac{-\sum x^2}{2(\theta_2-\theta_1^2)} + \frac{\theta_1 \sum x}{\theta_2-\theta_1^2} - \frac{n\theta_1^2}{2(\theta_2-\theta_1^2)} - \frac{n}{2}\log(\theta_2-\theta_1^2) - \frac{n}{2}\log 2\pi$$

Consider the expressions

$$\lambda_{11}\frac{\partial \log\Phi}{\partial\theta_1} + \lambda_{12}\frac{\partial\log\Phi}{\partial\theta_2} = \lambda_{11}\left[-\theta_1\frac{\sum x^2}{\theta_3^2} + \frac{\sum x}{\theta_3} + \frac{2\theta_1^2\sum x}{\theta_3^2} - \frac{n\theta_1^3}{\theta_3^2}\right]$$
$$+ \lambda_{12}\left[\frac{\sum x^2}{2\theta_3^2} - \frac{\theta_1\sum x}{\theta_3^2} + \frac{n\theta_1^2}{2\theta_3^2} - \frac{n}{2\theta_3}\right]$$

$$\lambda_{21}\frac{\partial \log\Phi}{\partial\theta_1} + \lambda_{22}\frac{\partial\log\Phi}{\partial\theta_2} = \lambda_{21}\left[-\theta_1\frac{\sum x^2}{\theta_3^2} + \frac{\sum x}{\theta_3} + \frac{2\theta_1^2\sum x}{\theta_3^2} - \frac{n\theta_1^3}{\theta_3^2}\right]$$
$$+ \lambda_{22}\left[\frac{\sum x^2}{2\theta_3^2} - \frac{\theta_1\sum x}{\theta_3^2} + \frac{n\theta_1^2}{2\theta_3^2} - \frac{n}{2\theta_3}\right]$$

where θ_3 has been written in place of $(\theta_2 - \theta_1^2)$

By choosing

$$\lambda_{11} = (\theta_2 - \theta_1^2)/n \quad ; \quad \lambda_{12} = 2\theta_1(\theta_2-\theta_1^2)/n$$
$$\lambda_{21} = 2\theta_1(\theta_2-\theta_1^2)/n \quad ; \quad \lambda_{22} = 2(\theta_2+\theta_1^2)(\theta_2-\theta_1^2)/n.$$

these forms reduce to

$$\left.\begin{array}{l}\sum x/n \quad - \quad \theta_1 \\ \sum x^2/n \quad - \quad \theta_2\end{array}\right\}$$

respectively. Consequently $\sum x/n$ is an unbiassed estimate of $\theta_1 = m$; $\sum x^2/n$ is an unbiassed estimate of $\theta_2 = m^2 + \sigma^2$, and these statistics make the generalised variance a minimum. The actual variance of $\sum x/n$, moreover, is

$$\lambda_{11} = \frac{\theta_2 - \theta_1^2}{n} = \frac{\sigma^2}{n}$$

That of $\sum x^2/n$ is

$$\lambda_{22} = 2(\theta_2 + \theta_1^2)(\theta_2 - \theta_1^2)/n = \frac{2\sigma^4}{n}\left(1 + 2\frac{m^2}{\sigma^2}\right)$$

The covariance of the pair of estimates is

$$\lambda_{12} = \lambda_{21} = 2\theta_1(\theta_2 - \theta_1^2)/n = 2m\sigma^2/n.$$

Hence the generalised variance is

$$\begin{vmatrix} \sigma^2/n & 2m\sigma^2/n \\ 2m\sigma^2/n & 2(\sigma^4/n)(1 + 2m^2/\sigma^2) \end{vmatrix} = 2\sigma^6/n^2.$$

5.2.1 Verification:- The last results can be checked by first principles. In a Bernouillian sample of n, the bivariate moment generating function of $\sum x/n$ and $\sum x^2/n$ is

$$M(\alpha \cdot \beta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\alpha \sum x/n + \beta \sum x^2/n} \; \Phi \, dx'$$

where $\Phi = (2\pi)^{-n/2} \cdot \sigma^{-n} \exp\left\{ - \sum_{i=1}^{n} (x_i - m)^2 / 2\sigma^2 \right\}$.

So

$$M(\alpha \cdot \beta) = \left[\frac{1}{\sqrt{2\pi} \cdot \sigma} \int_{-\infty}^{\infty} \exp\left\{ \frac{\alpha x}{n} + \frac{\beta x^2}{n} - \frac{x^2}{2\sigma^2} + \frac{m x}{\sigma^2} - \frac{m^2}{2\sigma^2} \right\} dx \right]^n$$

$$\therefore M(\alpha\cdot\beta) = \left[\left(\frac{n}{n-2\beta\sigma^2}\right)^{\frac{1}{2}} \exp\frac{\alpha^2\sigma^2+2mn(\alpha+m\beta)}{2n(n-2\beta\sigma^2)} \cdot \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \left(\frac{n-2\beta\sigma^2}{n}\right)^{\frac{1}{2}} \times \int_{-\infty}^{\infty} \exp\left\{-\left(x - \frac{mn+\alpha\sigma^2}{n-2\beta\sigma^2}\right)^2 \Big/ \frac{2\sigma^2 n}{n-2\beta\sigma^2}\right\} dx\right]^n$$

$$= \left(1 - \frac{2\beta\sigma^2}{n}\right)^{-\frac{n}{2}} \exp\left\{\frac{\alpha^2\sigma^2 + 2mn(\alpha+m\beta)}{2(n-2\beta\sigma^2)}\right\}$$

We are interested in the means and variances of the two statistics, i.e., in the linear and quadratic terms in the expansion of $M(\alpha\cdot\beta)$. Choosing $\beta < n/2\sigma^2$ - as we legitimately may - we find, after slight reduction,

$$M(\alpha\cdot\beta) = 1 + \alpha m + \beta(m^2+\sigma^2) + \frac{\alpha^2}{2!}\left(m^2 + \frac{\sigma^2}{n}\right)$$
$$+\frac{\beta^2}{2!}\left(\frac{n+2}{n}\sigma^4 + 2m^2\sigma^2 + \frac{4m^2\sigma^2}{n} + m^4\right) + \alpha\beta\left(m\sigma^2 + \frac{2m\sigma^2}{n} + m^3\right)$$
$$+ \cdots\cdots$$

From the theory of generating functions

(i) The mean of $\sum x/n$ is m.

(ii) The variance of $\sum x/n$ is $(m^2 + \sigma^2/n) - m^2 = \sigma^2/n$

(iii) The mean of $\sum x^2/n$ is $m^2 + \sigma^2$.

(iv) The variance of $\sum x^2/n$ is

$$\frac{n+2}{n}\sigma^4 + 2m^2\sigma^2 + \frac{4m^2\sigma^2}{n} + m^4 - (m^2+\sigma^2)^2$$

$$= \frac{2\sigma^4}{n}\left(1 + \frac{2m^2}{\sigma^2}\right).$$

(v) The covariance of $\sum x/n$ and $\sum x^2/n$ is

$$m\sigma^2 + \frac{2m\sigma^2}{n} + m^3 - m(m^2+\sigma^2)$$

$$= 2m\sigma^2/n.$$

These results agree completely with those already obtained. In particular, (i) and (iii) verify that $\sum x/n$, $\sum x^2/n$ are unbiassed estimates of m, $m^2+\sigma^2$ respectively.

It is not surprising that our present method of estimation does not admit $\sum(x-\bar{x})^2/n$ as a statistic, since this is biassed, its mean being not σ^2 but $\sigma^2\left(1-\frac{1}{n}\right)$. It is more unexpected at first sight that $\sum x/n$, $\sum(x-\bar{x})^2/\overline{n-1}$, which are both unbiassed, should not qualify. However, the sampling variance of the latter is found (by application of the moment generating function technique, or otherwise) to be $2\sigma^4/\overline{n-1}$. while the covariance with $\sum x/n$ is zero. Hence this pair has a generalised variance of

$$\begin{vmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/\overline{n-1} \end{vmatrix} = \frac{2\sigma^6}{n^2}\cdot\frac{n}{n-1}$$

which exceeds that of $\sum x/n$, $\sum x^2/n$ in the ratio $n/\overline{n-1}$.

The phenomenon of the basic parametric coefficients has been forcibly brought out in this example. Thus we were able to estimate m and $m^2+\sigma^2$ by means of unbiassed statistics of minimum generalised variance; but it was impossible to estimate m and σ^2 by such statistics.

5.3 Unbiassed Statistics of Minimum Generalised Variance and Koopman Distributions:- As we found in Section 5.1.4, two unbiassed statistics T_1, T_2 of minimum generalised variance satisfy

$$\left.\begin{aligned} \partial \log\Phi/\partial\theta_1 &= \alpha_{22}(T_1-\theta_1) - \alpha_{12}(T_2-\theta_2) \\ \partial \log\Phi/\partial\theta_2 &= -\alpha_{21}(T_1-\theta_1) + \alpha_{11}(T_2-\theta_2) \end{aligned}\right\} \qquad (14)$$

where $\alpha_{ij} = \lambda_{ij} / (\lambda_{11}\lambda_{22} - \lambda_{12}^2)$. Regarding the λ's (and hence the α's) and T's as given functions, we may solve this pair of simultaneous partial differential equations for Φ, and thus find the type of probability distribution which admits a pair of unbiassed statistics of minimum generalised variance.

Write

$$\int_0^{\theta_1} \alpha_{ij}\, d\theta_1 = {}_{(1)}F_{ij}(\theta_1 \cdot \theta_2) \quad ; \int_0^{\theta_2} \alpha_{ij}\, d\theta_2 = {}_{(2)}F_{ij}(\theta_1 \cdot \theta_2)$$

$$\int_0^{\theta_1} {}_{(1)}F_{ij}\, d\theta_1 = {}_{(1)}g_{ij}(\theta_1 \cdot \theta_2) \quad ; \int_0^{\theta_2} {}_{(2)}F_{ij}\, d\theta_2 = {}_{(2)}g_{ij}(\theta_1 \cdot \theta_2)$$

$[i, j = 1, 2]$

Integration of the first equation with respect to θ_1 gives $\log \Phi = (T_1 - \theta_1)\,{}_{(1)}F_{22} + {}_{(1)}g_{22} - (T_2 - \theta_2)\,{}_{(1)}F_{12} + C_1(\theta_2 \cdot x')$ where C_1 is a constant of integration, independent of θ_1.

Similarly, the integral of the second equation is

$$\log \Phi = -(T_1 - \theta_1)\,{}_{(2)}F_{12} + (T_2 - \theta_2)\,{}_{(2)}F_{11} + {}_{(2)}g_{11} + C_2(\theta_1 \cdot x')$$

where C_2 does not involve θ_2.

Adding

$$\log \Phi = \frac{1}{2}\Big[(T_1 - \theta_1)({}_{(1)}F_{22} - {}_{(2)}F_{12}) + (T_2 - \theta_2)({}_{(2)}F_{11} - {}_{(1)}F_{12}) + {}_{(1)}g_{22} + {}_{(2)}g_{11} + C(\theta_1 \cdot \theta_2 \cdot x')\Big]$$

where $C(\theta_1 \cdot \theta_2 \cdot x') = C_1(\theta_2 \cdot x') + C_2(\theta_1 \cdot x')$.

The necessary and sufficient condition that this be the solution of the original pair of differential equations is $C(\theta_1 \cdot \theta_2 \cdot x')$ = function of x' alone. Differentiating the proposed form of $\log \Phi$ we obtain

$$\frac{\partial \log \Phi}{\partial \theta_1} = \frac{1}{2}\left[\alpha_{22}(T_1 - \theta_1) - \alpha_{12}(T_2 - \theta_2) - (T_1 - \theta_1)\int \frac{\partial \alpha_{12}}{\partial \theta_1} d\theta_2 + (T_2 - \theta_2)\int \frac{\partial \alpha_{11}}{\partial \theta_1} d\theta_2 + {}_{(2)}F_{11} + {}_{(2)}F_{12} + \frac{\partial C}{\partial \theta_1}\right] \qquad (15)$$

From (14)

$$\frac{\partial}{\partial\theta_2}\left\{\alpha_{22}(T_1-\theta_1)-\alpha_{12}(T_2-\theta_2)\right\}=\frac{\partial}{\partial\theta_1}\left\{-\alpha_{21}(T_1-\theta_1)+\alpha_{11}(T_2-\theta_2)\right\}$$

whence, integrating both sides with respect to θ_2,

$$\alpha_{22}(T_1-\theta_1)-\alpha_{12}(T_2-\theta_2)=-(T_1-\theta_1)\int\frac{\partial\alpha_{21}}{\partial\theta_1}d\theta_2+(T_2-\theta_2)\int\frac{\partial\alpha_{11}}{\partial\theta_1}d\theta_2 + {}_{(2)}F_{11}+{}_{(2)}F_{12}$$

Substituting in (15),

$$\frac{\partial\log\Phi}{\partial\theta_1}=\alpha_{22}(T_1-\theta_1)-\alpha_{12}(T_2-\theta_2)+\frac{1}{2}\frac{\partial C}{\partial\theta_1}$$

which agrees with the first equation of (14) provided $\partial C/\partial\theta_1=0$; i.e provided C is independent of θ_1. Similarly, to satisfy the second equation of (14), C must be independent of θ_2. Thus C is necessarily a function of x' only; and this condition is clearly sufficient. A distribution admitting two unbiassed statistics of minimum generalised variance has therefore the form

$$\log\Phi=\frac{1}{2}\left[(T_1-\theta_1)({}_{(1)}F_{22}-{}_{(2)}F_{12})+(T_2-\theta_2)({}_{(2)}F_{11}-{}_{(1)}F_{12})+{}_{(1)}g_{22}+{}_{(2)}g_{11}+C(x')\right]$$

or, writing

$$F_1(\theta_1\cdot\theta_2)=\tfrac{1}{2}\{{}_{(1)}F_{22}-{}_{(2)}F_{12}\}$$

$$F_2(\theta_1\cdot\theta_2)=\tfrac{1}{2}\{{}_{(2)}F_{11}-{}_{(1)}F_{12}\}$$

$$n\,F_3(\theta_1\cdot\theta_2)=\tfrac{1}{2}\{{}_{(1)}g_{22}+{}_{(2)}g_{11}-2\theta_1F_1(\theta_1\cdot\theta_2)-2\theta_2F_2(\theta_1\cdot\theta_2)\}$$

$$\log\Phi=T_1(x')F_1(\theta_1\cdot\theta_2)+T_2(x')F_2(\theta_1\cdot\theta_2)+nF_3(\theta_1\cdot\theta_2)+\tfrac{1}{2}C(x').$$

Since $\log\Phi$ is the sum of n similar terms $\sum_j\log\varphi(x_j|\theta_1\cdot\theta_2)$ there exist functions $f_1(x), f_2(x), c(x)$ such that

$$\sum_{j=1}^{n} f_i(x_j)=T_i(x')\quad[i=1,2]$$

$$\sum_{j=1}^{n} c(x_j)=\tfrac{1}{2}C(x').$$

whence, finally,

$$\varphi = \exp\{f_1(x)F_1(\theta_1\cdot\theta_2) + f_2(x)F_2(\theta_1\cdot\theta_2) + F_3(\theta_1\cdot\theta_2) + c(x)\},$$

a distribution of Koopman's form, admitting two sufficient statistics. By Koopman's Second Theorem, T_1, T_2 are themselves sufficient.

In (14), it was assumed that T_1, T_2 and the λ's were all prescribed. The only unknown element in φ is therefore $c(x)$ which must, however, be such that φ obeys the total probability condition over a given range. We shall study the implications of this statement in Section 5.4.

<u>5.3.1</u> <u>The Converse Problem</u> to the foregoing is - given the general Koopman two-parameter distribution, does it admit a pair of unbiassed statistics of minimum generalised variance? If so, what are they and what coefficients do they estimate?

For the "canonical" form of Koopman distribution

$$\varphi = \exp\{\psi_1 f_1(x) + \psi_2 f_2(x) + F(\psi_1\cdot\psi_2) + c(x)\}$$

we have, in a Bernouillian sample of n,

$$\log \Phi = nT_1\psi_1 + nT_2\psi_2 + nF + \sum_{j=1}^{n} c(x_j)$$

where T_i has been written for

$$\frac{1}{n}\sum_{j=1}^{n} f_i(x_j) \qquad [i = 1, 2].$$

Suppose that statistics u_1, u_2 exist, unbiassed estimates of θ_1, θ_2 respectively, which minimise the generalised variance. This implies that functions λ_{11}, $\lambda_{12} = \lambda_{21}$, λ_{22} (of ψ_1 and ψ_2) can be found such that

$$\mu_i - \theta_i \equiv \frac{1}{\Phi}\left(\lambda_{i1}\frac{\partial \Phi}{\partial \theta_1} + \lambda_{i2}\frac{\partial \Phi}{\partial \theta_2}\right) \qquad [i = 1, 2]$$

i.e., such that

$$\mu_1 - \theta_1 \equiv nT_1\left(\lambda_{11}\frac{\partial \psi_1}{\partial \theta_1} + \lambda_{12}\frac{\partial \psi_1}{\partial \theta_2}\right) + nT_2\left(\lambda_{11}\frac{\partial \psi_2}{\partial \theta_1} + \lambda_{12}\frac{\partial \psi_2}{\partial \theta_2}\right) + n\left(\lambda_{11}\frac{\partial F}{\partial \theta_1} + \lambda_{12}\frac{\partial F}{\partial \theta_2}\right)$$

$$\mu_2 - \theta_2 \equiv nT_1\left(\lambda_{21}\frac{\partial \psi_1}{\partial \theta_1} + \lambda_{22}\frac{\partial \psi_1}{\partial \theta_2}\right) + nT_2\left(\lambda_{21}\frac{\partial \psi_2}{\partial \theta_1} + \lambda_{22}\frac{\partial \psi_2}{\partial \theta_2}\right) + n\left(\lambda_{21}\frac{\partial F}{\partial \theta_1} + \lambda_{22}\frac{\partial F}{\partial \theta_2}\right)$$

Consider the former identity. In order that it be of the requisite form, it is necessary that the coefficients of T_1, T_2 should be constants, i.e.,

$$\left.\begin{aligned} \lambda_{11}\frac{\partial \psi_1}{\partial \theta_1} + \lambda_{12}\frac{\partial \psi_1}{\partial \theta_2} &= k_1/n \\ \lambda_{11}\frac{\partial \psi_2}{\partial \theta_1} + \lambda_{12}\frac{\partial \psi_2}{\partial \theta_2} &= k_2/n \end{aligned}\right\}$$

while it is also necessary that

$$\lambda_{11}\frac{\partial F}{\partial \theta_1} + \lambda_{12}\frac{\partial F}{\partial \theta_2} = -\theta_1/n$$

Solving the first two of these necessary equations,

$$\lambda_{11} = \left(k_1\frac{\partial \psi_2}{\partial \theta_2} - k_2\frac{\partial \psi_1}{\partial \theta_2}\right)\Big/ nJ \tag{16}$$

$$\lambda_{12} = \left(k_2\frac{\partial \psi_1}{\partial \theta_1} - k_1\frac{\partial \psi_2}{\partial \theta_1}\right)\Big/ nJ \tag{17}$$

where $J = \frac{\partial \psi_1}{\partial \theta_1}\cdot\frac{\partial \psi_2}{\partial \theta_2} - \frac{\partial \psi_1}{\partial \theta_2}\cdot\frac{\partial \psi_2}{\partial \theta_1} = \frac{\partial(\psi_1\cdot\psi_2)}{\partial(\theta_1\cdot\theta_2)}$ in the usual Jacobian notation. Substituting these values in the third necessary equation,

$$k_1\left(\frac{\partial F}{\partial \theta_1}\frac{\partial \psi_2}{\partial \theta_2} - \frac{\partial F}{\partial \theta_2}\frac{\partial \psi_2}{\partial \theta_1}\right) + k_2\left(\frac{\partial F}{\partial \theta_2}\frac{\partial \psi_1}{\partial \theta_1} - \frac{\partial F}{\partial \theta_1}\frac{\partial \psi_1}{\partial \theta_2}\right) = -\theta_1 J$$

or $$k_1\frac{\partial(F\cdot\psi_2)}{\partial(\theta_1\cdot\theta_2)} + k_2\frac{\partial(\psi_1\cdot F)}{\partial(\theta_1\cdot\theta_2)} = -\theta_1\frac{\partial(\psi_1\cdot\psi_2)}{\partial(\theta_1\cdot\theta_2)}$$

or, utilising the property of Jacobians

$$\frac{\partial(u\cdot v)}{\partial(x\cdot y)} \div \frac{\partial(s\cdot t)}{\partial(x\cdot y)} = \frac{\partial(u\cdot v)}{\partial(s\cdot t)}$$

$$k_1 \frac{\partial(F \cdot \psi_2)}{\partial(\psi_1 \cdot \psi_2)} + k_2 \frac{\partial(\psi_1 \cdot F)}{\partial(\psi_1 \cdot \psi_2)} = -\theta_1$$

That is, on expanding the Jacobians,

$$k_1 \frac{\partial F}{\partial \psi_1} + k_2 \frac{\partial F}{\partial \psi_2} = -\theta_1 \tag{18}$$

In other words, the parameter θ_1 is a linear combination of $-\frac{\partial F}{\partial \psi_1}$ and $-\frac{\partial F}{\partial \psi_2}$. Similarly, the second parameter θ_2 is a linear combination of these same basic coefficients, say

$$-\theta_2 = k_3 \frac{\partial F}{\partial \psi_1} + k_4 \frac{\partial F}{\partial \psi_2} \tag{19}$$

corresponding to which we have

$$\lambda_{21} = \left(k_3 \frac{\partial \psi_2}{\partial \theta_2} - k_4 \frac{\partial \psi_1}{\partial \theta_2}\right) \Big/ nJ \tag{20}$$

$$\lambda_{22} = \left(k_4 \frac{\partial \psi_1}{\partial \theta_1} - k_3 \frac{\partial \psi_2}{\partial \theta_1}\right) \Big/ nJ \tag{21}$$

For consistency, we must have $\lambda_{12} = \lambda_{21}$, that is, we require

$$k_2 \frac{\partial \psi_1}{\partial \theta_1} - k_1 \frac{\partial \psi_2}{\partial \theta_1} = k_3 \frac{\partial \psi_2}{\partial \theta_2} - k_4 \frac{\partial \psi_1}{\partial \theta_2}$$

Differentiate both (18) and (19) with respect to θ_1 and θ_2;

$$k_1\left(\frac{\partial^2 F}{\partial \psi_1^2}\frac{\partial \psi_1}{\partial \theta_1} + \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_2}{\partial \theta_1}\right) + k_2\left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_1}{\partial \theta_1} + \frac{\partial^2 F}{\partial \psi_2^2}\frac{\partial \psi_2}{\partial \theta_1}\right) = -1$$

$$k_1\left(\frac{\partial^2 F}{\partial \psi_1^2}\frac{\partial \psi_1}{\partial \theta_2} + \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_2}{\partial \theta_2}\right) + k_2\left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_1}{\partial \theta_2} + \frac{\partial^2 F}{\partial \psi_2^2}\frac{\partial \psi_2}{\partial \theta_2}\right) = 0$$

$$k_3\left(\frac{\partial^2 F}{\partial \psi_1^2}\frac{\partial \psi_1}{\partial \theta_1} + \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_2}{\partial \theta_1}\right) + k_4\left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_1}{\partial \theta_1} + \frac{\partial^2 F}{\partial \psi_2^2}\frac{\partial \psi_2}{\partial \theta_1}\right) = 0$$

$$k_3\left(\frac{\partial^2 F}{\partial \psi_1^2}\frac{\partial \psi_1}{\partial \theta_2} + \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_2}{\partial \theta_2}\right) + k_4\left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\frac{\partial \psi_1}{\partial \theta_2} + \frac{\partial^2 F}{\partial \psi_2^2}\frac{\partial \psi_2}{\partial \theta_2}\right) = -1$$

From the first and third of these, we obtain $\partial\psi_1/\partial\theta_1$ and $\partial\psi_2/\partial\theta_1$; while $\partial\psi_1/\partial\theta_2$, $\partial\psi_2/\partial\theta_2$ come from the second and fourth. The values are

$$\frac{\partial\psi_1}{\partial\theta_1}=\left\{-\left(k_3\frac{\partial^2F}{\partial\psi_1\partial\psi_2}+k_4\frac{\partial^2F}{\partial\psi_2^2}\right)\Big/D\right\};\ \frac{\partial\psi_2}{\partial\theta_1}=\left(k_3\frac{\partial^2F}{\partial\psi_1^2}+k_4\frac{\partial^2F}{\partial\psi_1\partial\psi_2}\right)\Big/D$$

$$\frac{\partial\psi_1}{\partial\theta_2}=\left\{\left(k_1\frac{\partial^2F}{\partial\psi_1\partial\psi_2}+k_2\frac{\partial^2F}{\partial\psi_2^2}\right)\Big/D\right\};\ \frac{\partial\psi_2}{\partial\theta_2}=-\left(k_1\frac{\partial^2F}{\partial\psi_1^2}+k_2\frac{\partial^2F}{\partial\psi_1\partial\psi_2}\right)\Big/D$$

where

$$D=\left(k_1\,\partial^2F/\partial\psi_1^2+k_2\,\partial^2F/\partial\psi_1\partial\psi_2\right)\left(k_3\,\partial^2F/\partial\psi_1\partial\psi_2+k_4\,\partial^2F/\partial\psi_2^2\right)$$
$$-\left(k_1\,\partial^2F/\partial\psi_1\partial\psi_2+k_2\,\partial^2F/\partial\psi_2^2\right)\left(k_3\,\partial^2F/\partial\psi_1^2+k_4\,\partial^2F/\partial\psi_1\partial\psi_2\right)$$

Substitution of these values in (17) and (20) confirms that

$$\lambda_{12}=\lambda_{21}$$

as required. Consequently, the general Koopman two-parameter distribution does admit pairs of unbiassed statistics of minimum generalised variance. All these pairs are, however, linear combinations of one fundamental, or "basic" pair. Thus the statistics

$$u_1\equiv k_1\left\{\frac{1}{n}\sum_{j=1}^{n}f_1(x_j)\right\}+k_2\left\{\frac{1}{n}\sum_{j=1}^{n}f_2(x_j)\right\}$$

$$u_2\equiv k_3\left\{\frac{1}{n}\sum_{j=1}^{n}f_1(x_j)\right\}+k_4\left\{\frac{1}{n}\sum_{j=1}^{n}f_2(x_j)\right\}$$

(the k's being constants) are unbiassed estimates of

$$-k_1\frac{\partial F}{\partial\psi_1}\quad-k_2\frac{\partial F}{\partial\psi_2}$$

$$-k_3\frac{\partial F}{\partial\psi_1}\quad-k_4\frac{\partial F}{\partial\psi_2}$$

respectively, and give minimum generalised variance. The coefficients $-\partial F/\partial\psi_1$, $-\partial F/\partial\psi_2$ will be termed the basic parametric coefficients of the distribution.

By inserting the values of $\partial\psi_i/\partial\theta_j$ $[i,j=1,2]$ in (16), (17), (20), (21) we obtain explicitly the variances and covariance:

(i) Variance of $\mathcal{U}_1$ is (equations (10) and (16))

$$-\frac{1}{n}\left[k_1^2\frac{\partial^2 F}{\partial\psi_1^2}+2k_1k_2\frac{\partial^2 F}{\partial\psi_1\partial\psi_2}+k_2^2\frac{\partial^2 F}{\partial\psi_2^2}\right]$$

(ii) Variance of $\mathcal{U}_2$ is, similarly,

$$-\frac{1}{n}\left[k_3^2\frac{\partial^2 F}{\partial\psi_1^2}+2k_3k_4\frac{\partial^2 F}{\partial\psi_1\partial\psi_2}+k_4^2\frac{\partial^2 F}{\partial\psi_2^2}\right]$$

(iii) Covariance of $\mathcal{U}_1$ and $\mathcal{U}_2$ is (equation (12) and (17))

$$-\frac{1}{n}\left[k_1k_3\frac{\partial^2 F}{\partial\psi_1^2}+(k_2k_3+k_1k_4)\frac{\partial^2 F}{\partial\psi_1\partial\psi_2}+k_2k_4\frac{\partial^2 F}{\partial\psi_2^2}\right]$$

The generalised variance is therefore

$$n^{-2}\cdot\begin{vmatrix}k_1 & k_2\\ k_3 & k_4\end{vmatrix}^2\cdot\begin{vmatrix}\partial^2F/\partial\psi_1^2 & \partial^2F/\partial\psi_2\partial\psi_1\\ \partial^2F/\partial\psi_1\partial\psi_2 & \partial^2F/\partial\psi_2^2\end{vmatrix}$$

<u>5.3.2</u> <u>A Special Case</u>- A case of special importance occurs when we choose $k_1=k_4=1;\ k_2=k_3=0$. The foregoing results then reduce to the following:

$T_1=\frac{1}{n}\sum_{j=1}^{n}f_1(x_j)$ is an unbiassed estimate of $-\dfrac{\partial F}{\partial\psi_1}$.

$T_2=\frac{1}{n}\sum_{j=1}^{n}f_2(x_j)$ is an unbiassed estimate of $-\dfrac{\partial F}{\partial\psi_2}$.

The variances of T_1, T_2 respectively are

$$-\frac{1}{n}\frac{\partial^2 F}{\partial\psi_1^2}\quad\text{and}\quad-\frac{1}{n}\frac{\partial^2 F}{\partial\psi_2^2}$$

and their covariance is $-\frac{1}{n}\frac{\partial^2 F}{\partial\psi_1\partial\psi_2}$. The generalised variance (which is a minimum) is

$$-n^{-2}\begin{vmatrix} \partial^2 F/\partial\psi_1^2 & \partial^2 F/\partial\psi_2\partial\psi_1 \\ \partial^2 F/\partial\psi_1\partial\psi_2 & \partial^2 F/\partial\psi_2^2 \end{vmatrix}$$

These expressions justify the assertion, made in 5.1.1, that tl the variances and covariance in large samples of n are each $O(n^{-1})$.

5.3.3 Examples of Basic Parametric Coefficients:-

(a) The normal distribution

$$\varphi = \frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$$

can be transformed, on putting $\psi_1 = 1/2\sigma^2$, $\psi_2 = m/\sigma^2$ into the Koopman canonical form

$$\varphi = \exp\left\{-\psi_1 x^2 + \psi_2 x - \frac{\psi_2^2}{4\psi_1} + \frac{1}{2}\log\psi_1 - \frac{1}{2}\log 2\pi\right\}.$$

The basic parametric coefficients are accordingly

(i) $-\frac{\partial}{\partial\psi_1}\left(-\frac{\psi_2^2}{4\psi_1} + \frac{1}{2}\log\psi_1\right) = -\frac{\psi_2^2}{4\psi_1^2} - \frac{1}{2\psi_1} = -(m^2+\sigma^2)$

which is estimated by $-\frac{1}{n}\sum_{j=1}^{n} x_j^2$.

(ii) $-\frac{\partial}{\partial\psi_2}\left(-\frac{\psi_2^2}{4\psi_1} + \frac{1}{2}\log\psi_1\right) = \frac{\psi_2}{2\psi_1} = m$,

which is estimated by $\frac{1}{n}\sum_{j=1}^{n} x_j$.

These results confirm those in Section 5.2.

(b) The Pearson curve

$$\varphi = \frac{1}{a\Gamma(p+1)}\left(\frac{x}{a}\right)^p e^{-x/a} \qquad (x \geq 0)$$

can be transformed, on putting $a = 1/a'$ into the Koopman canonical form

$$\varphi = \exp\{-a'x + p\log x + (p+1)\log a' - \log\Gamma(p+1)\}.$$

The basic parametric coefficients are accordingly

$$\text{(i)} -\frac{\partial}{\partial a'}\left\{(p+1)\log a' - \log\Gamma(p+1)\right\} = -\frac{p+1}{a'}$$

which is estimated by $-\frac{1}{n}\sum_{j=1}^{n} x_j$

$$\text{(ii)} -\frac{\partial}{\partial p}\left\{(p+1)\log a' - \log\Gamma(p+1)\right\} = -\log a' + \left\{\frac{d\,\Gamma(p+1)}{d(p+1)}\right\}\frac{1}{\Gamma(p+1)}$$

which is estimated by $\frac{1}{n}\sum_{j=1}^{n}\log x_j$.

We note that for this distribution the unbiassed statistics of minimum generalised variance are the arithmetic and geometric means.

5.4 <u>Unbiassed Statistics of Minimum Generalised Variance and the Total Probability Conditioh</u>:- In Section 5.3 we learned that the question "is there a distribution, over a given range, (a, b) for which two prescribed statistics, with prescribed variances and covariance, are unbiassed and of minimum generalised variance?" is synonymous with "can a function $c(x)$ be determined such that

$$\int_a^b \exp\{\psi_1 f_1(x) + \psi_2 f_2(x) + F(\psi_1 \cdot \psi_2) + c(x)\}\,dx = 1$$

for all values of ψ_1, ψ_2; the f's and F being given?"

Writing $\exp c(x) = \mu(x)$ this equation is equivalent to

$$\int_a^b \mu(x)\exp\{\psi_1 f_1(x) + \psi_2 f_2(x)\}\,dx = \exp\{-F(\psi_1 \cdot \psi_2)\}.$$

Now there is in general <u>no</u> function $\mu(x)$ for which this relation is true, as we can show by the "passage to the limit" technique of the Calculus of Functionals. Thus, subdivide the range of x into t equal parts by the values

$$x_1 = a,\ x_2,\ \cdots\cdots\ x_i,\ \cdots\cdots\ x_{t+1} = b$$

$$[x_i - x_{i-1} = h, \text{ say}]$$

Let the ranges of values of ψ_1 and ψ_2 be also divided into t parts by the values

$$\psi_{j(1)} = \alpha_j,\ \psi_{j(2)},\ \cdots\cdots\ \psi_{j(i)},\ \cdots\ \psi_{j(t+1)} = \beta_j \quad [j = 1, 2]$$

The integral equation above is the limit, as $t \to \infty$, $h \to 0$, $th \to b - a$, of the system of linear algebraic equations

$$\sum_{i=1}^{t+1} u(x_i)\, h\left[\exp\{\psi_{1(j)} f_1(x_i) + \psi_{2(k)} f_2(x_i)\}\right] = e^{-F(\psi_{1(j)},\, \psi_{2(k)})}.$$

$$[j = 1, 2, \cdots t+1; \quad k = 1, 2, \cdots t+1]$$

This system consists of $(t+1)^2$ equations in $t+1$ unknowns $u(x_1), \cdots u(x_i) \cdots u(x_{t+1})$. There is therefore no set of values of the unknowns which satisfies all the equations, in general. Passing to the limit $(t \to \infty)$ it follows that in general no function $u(x)$ makes

$$\int_a^b u(x) \exp\{\psi_1 f_1(x) + \psi_2 f_2(x)\}\, dx = e^{-F(\psi_1, \psi_2)}$$

for all values of ψ_1, ψ_2.

In certain special cases, a solution may exist. Let us ascertain the relevant circumstances, and the corresponding solution. Supposing for the moment that the limits are $\pm\infty$, we try to find $u(x)$ such that

$$\int_{-\infty}^{\infty} u(x) \exp\{\psi_1 f_1(x) + \psi_2 f_2(x)\}\, dx = e^{-F(\psi_1, \psi_2)}$$

for all values of ψ_1, ψ_2. Put $f_1(x) = iy$, and denote the inverse transformation by $x = -i f_1^{-1}(y)$ If the limits of y, also, are $\pm\infty$ we have

$$\int_{-\infty}^{\infty} u[-if_1^{-1}(y)] \exp\{i\psi_1 y + \psi_2 f_2[-if_1^{-1}(y)]\}[-i\frac{df_1^{-1}(y)}{dy}]\,dy$$

$$\equiv e^{-F(\psi_1 \cdot \psi_2)} \text{ for all values of } \psi_1, \psi_2.$$

or, writing $-iu[-if_1^{-1}(y)]\frac{df_1^{-1}(y)}{dy} = u^x(y)$; $f_2[-if_1^{-1}(y)] = f_2^x(y)$,

$$\int_{-\infty}^{\infty} u^x(y) \exp\{i\psi_1 y + \psi_2 f_2^x(y)\}\,dy = e^{-F(\psi_1 \cdot \psi_2)} \text{ for all } \psi_1, \psi_2.$$

In particular, when $\psi_2 = 0$,

$$\int_{-\infty}^{\infty} u^x(y) \exp(i\psi_1 y)\,dy = e^{-F(\psi_1 \cdot 0)}$$

Hence, assuming F is bounded, the theory of Laplace transforms is applicable, and

$$u^x(y) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \exp\{-i\psi_1 y - F(\psi_1 \cdot 0)\}\,d\psi_1 .$$

By hypothesis, this value of $u^x(y)$ must satisfy the original equation for <u>all</u> values of ψ_2, not merely for $\psi_2 = 0$. Accordingly,

$$\int_{y=-\infty}^{\infty} \exp\{i\psi_1 y + \psi_2 f_2^x(y)\}\,dy \int_{t=-\infty}^{\infty} \exp\{-ity - F(t\cdot 0)\}\,dt \equiv 2\pi i e^{-F(\psi_1 \cdot \psi_2)}$$

or, utilising the transformation $x = -if_1^{-1}(y)$,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{df_1(x)}{dx} \exp\{(\psi_1 - t) f_1(x) + \psi_2 f_2(x) - F(t\cdot 0)\}\,dt\,dx \equiv 2\pi e^{-F(\psi_1 \cdot \psi_2)} \quad (22)$$

If, then, f_1, f_2, F satisfy this condition - which is both necessary and sufficient - a probability distribution does exist for which $\sum \frac{1}{n} f_1$, $\sum \frac{1}{n} f_2$ are unbiassed statistics of minimum generalised variance. Since $u^x(y)$ or

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty} \exp\{-i\psi_1 y - F(\psi_1 \cdot 0)\}\,d\psi_1$$

is determined, we can determine the hitherto unknown element $c(x)$, whereupon the distribution is completely specified. By the uniqueness property of Laplace transforms there is at most one Koopman distribution corresponding to functions $f_1(x), f_2(x), F(\psi_1 \cdot \psi_2)$.

If the range of the variate is not $\mp\infty$ (or if the transformations employed do not preserve the range), the foregoing general statements remain true; but instead of employing Laplace transforms, we must invert the integrals by the general theory of integral equations.

5.4.1 The form of F:- The total probability condition for our distribution may be put in the form

$$e^{-F(\psi_1 \cdot \psi_2)} = \int exp\{\psi_1 f_1(x) + \psi_2 f_2(x) + c(x)\} dx \qquad (23)$$

Thus, given merely $f_1(x)$, $f_2(x)$ there is an infinite number of Koopman distributions over a prescribed range - one to each function $c(x)$, which may be chosen arbitrarily. Equation (23) should, in fact, be considered as defining the class of functions F for which two-parameter probability distributions, admitting sufficient statistics, exist. When (23) is satisfied, (22) is also valid; the dintinction between these two formulae is that the latter contains no explicit mention of $c(x)$, which we originally encountered as an arbitrary "constant" of integration.

5.4.2 The Non-Existence of a Pair of Unbiassed Statistics of Minimum Variance:- In Chapter Four we deferred consideration of the question whether a distribution exists which admits two unbiassed statistics of minimum variance. Equation (23) now presents us with the means of tackling the problem.

We recall that, for unbiassed statistics of minimum variance, the $F(\psi_1 \cdot \psi_2)$ of Koopman's distribution degenerated into the form

$$U(\psi_1) + V(\psi_2)$$

(U, V are arbitrary functions of one variable). Corresponding to this F, we must try to determine a function $c(x)$ such that (23) is observed, i.e., such that

$$-U(\psi_1) - V(\psi_2) = \log \int \exp\{\psi_1 f_1(x) + \psi_2 f_2(x) + c(x)\}\, dx .$$

If a $c(x)$ exists, we have (on differentiating partially with respect to ψ_1 and then ψ_2)

$$\iint e^{c(x)+c(y)} \left[\begin{array}{c} e^{\psi_1\{f_1(x)+f_1(y)\} + \psi_2\{f_2(x)+f_2(y)\}} \\ \times f_1(x)\{f_2(x) - f_2(y)\} \end{array} \right] dx\, dy = 0$$

for all values of ψ_1 and ψ_2.

Regard the terms within the bracket [ ] as the kernel K of an integral equation,

$$\iint \omega(x \cdot y)\, K(x \cdot y \cdot \psi_1 \cdot \psi_2)\, dx\, dy = 0$$

This equation has only one continuous solution, $\omega \equiv 0$. Therefore, when ω is replaced by $\exp\{c(x) + c(y)\}$ which is positive for finite c, the foregoing double integral is not identically zero, unless indeed the kernel itself is. The latter occurs only if

$f_2(x) = f_2(y)$ for all values of x and y;

i.e., if $f_2(x) =$ constant, say α.

The distribution consequently reduces to

$$\varphi = \exp\left\{ \psi_1 f_1(x) + \alpha \psi_2 + \mathcal{U}(\psi_1) + V(\psi_2) + c(x) \right\}$$

By the total probability condition,

$$\left[\exp\{\alpha\psi_2 + V(\psi_2)\}\right]\left[\int \exp\{\psi_1 f_1(x) + \mathcal{U}(\psi_1) + c(x)\}\,dx\right] = 1$$

That is,

(a function of ψ_2) $\times$ (a function of ψ_1) $= 1$,

for all values of ψ_1, ψ_2. Each factor is accordingly a constant, and

$$\alpha\psi_2 + V(\psi_2) = 0.$$

There is therefore no probability distribution which admits two unbiassed statistics of minimum variance, and the criteria of Chapter Four are completely sterile.

5.4.3 A Property of the Arithmetic Mean. We have already shown that the only one-parameter distribution, over $\pm\infty$ for which the mean, with variance independent of the coefficient estimated, is sufficient, is the normal curve. The two-parameter analogue will now be considered.

If two sufficient statistics exist, the parent distribution must be of Koopman's form; if one of the statistics is the arithmetic mean, this form is

$$\varphi = \exp\{\psi_1 x + \psi_2 f_2(x) + F(\psi_1, \psi_2) + c(x)\}.$$

By Section 5.4.1 we can choose $f_2(x)$, $c(x)$, and the range, arbitrarily, and determine F such that φ obeys the total probability condition. Therefore there is an infinite number of distributions for which the mean is sufficient.

Suppose we investigate now the distributions for which the mean is a sufficient and unbiassed statistic, with variance of given form. Thus let $\frac{1}{n}\sum x$ be an unbiassed estimate of a coefficient $-\frac{\partial F}{\partial \psi_1}$ with variance of, say, $1/2n\psi_2$.

We have, for the F of the Koopman distribution sought for

$$\frac{\partial^2 F}{\partial \psi_1^2} = -\frac{1}{2\psi_2}$$

$$\therefore \quad F = -\frac{\psi_1^2}{4\psi_2} + g(\psi_2) + \psi_1 h(\psi_2)$$

where g, h are arbitrary functions. The distribution itself is

$$\varphi = \exp\left\{\psi_1 x + \psi_2 f_2(x) - \frac{\psi_1^2}{4\psi_2} + g(\psi_2) + \psi_1 h(\psi_2) + c(x)\right\}$$

and the total probability condition demands (for a doubly-infinite range of x) that

$$\int_{-\infty}^{\infty} \exp\{\psi_1 x + \psi_2 f_2(x) + c(x)\}\,dx = \exp\left\{\frac{\psi_1^2}{4\psi_2} - g(\psi_2) - \psi_1 h(\psi_2)\right\} \quad (24)$$

for all values of ψ_1, ψ_2. Giving ψ_2 the particular value 1 and writing

$$A = \exp\{-g(1)\} \; ; \quad B = h(1)$$

we can solve

$$\int_{-\infty}^{\infty} \exp\{\psi_1 x + f_2(x) + c(x)\}\,dx = A \exp\left\{\frac{\psi_1^2}{4} - B\psi_1\right\}$$

for the unknown $\exp\{f_2(x) + c(x)\}$, and the result obtained is unique. Thus, corresponding to any $f_2(x)$, we determine one function of x say $c_1(x)$, which satisfies the last equation. Moreover, this function $c_1(x)$ must also satisfy (24) for all values of ψ_1, ψ_2;

$$\int_{-\infty}^{\infty} \exp\{\psi_1 x + \psi_2 f_2(x) + c_1(x)\}\,dx = \exp\left\{\frac{\psi_1^2}{4\psi_2} - g(\psi_2) - \psi_1 h(\psi_2)\right\}.$$

Since $c_1(x)$ is known in terms of $f_2(x)$ and the constants A, B, this represents a (non-linear) integral equation in a function of one variable, which must hold for all values of two other variables ψ_1, ψ_2. By the method of Section 5.4, we can show that such an equation has in general no solution; but that if, in a particular case, a solution does exist, it is unique. Now in the present instance we know a solution of (24) viz., $f_2(x) = -x^2$; $c_1(x) = 0$, corresponding to which (equation (23)

$$e^{-F(\psi_1, \psi_2)} = \int_{-\infty}^{\infty} \exp\{\psi_1 x - \psi_2 x^2\}\, dx = \sqrt{\frac{\pi}{\psi_2}} \cdot e^{\psi_1^2/4\psi_2}$$

$$\therefore F(\psi_1, \psi_2) = -\frac{\psi_1^2}{4\psi_2} + \frac{1}{2}\log \psi_2 - \frac{1}{2}\log \pi.$$

(i.e., in (24), $g \equiv \frac{1}{2}\log\psi_2 - \frac{1}{2}\log\pi$; $h \equiv 0$)

The Koopman distribution possessing these elements is the normal distribution (Section (5.3.3). Since, as we have seen, no other expressions for f_2, c satisfy (24), we have the theorem:

The only distribution, with range $\pm\infty$, for which the mean is a sufficient and unbiassed estimate of $\psi_1/2\psi_2$ with variance $1/2n\psi_2$, is the normal curve.

A similar theorem has been enunciated by R. A. Fisher - "the only Pearsonian distribution for which the mean is a sufficient statistic of location is the normal distribution." At the cost of particularising both the coefficient to be estimated and the variance of the estimate, we have been able to remove the adjective "Pearsonian."

In Section 5.3.3, we encountered a second Pearson curve - the Gamma Type - for which the arithmetic mean was a sufficient statistic. It may be of interest, therefore, to carry out a census of this family.

5.4.4 Sufficient Statistics Admitted by Pearson Curves:- The members of Pearson's family of distributions, defined by

$$\frac{d\varphi}{dx} = -\frac{(x-a)\varphi}{c_0 + c_1 x + c_2 x^2}$$

are classified according to the nature of the coefficients of the denomimator.

(A) Denominator possessing real, unequal roots, say α_1, α_2.

We have

$$\log \varphi = -\frac{1}{c_2}\int \frac{(x-a)\,dx}{(x-\alpha_1)(x-\alpha_2)}$$

$$= -\frac{\alpha_1 - a}{c_2(\alpha_1-\alpha_2)}\log(x-\alpha_1) + \frac{\alpha_2 - a}{c_2(\alpha_1-\alpha_2)}\log(x-\alpha_2) + A$$

where A is a constant. Since $\int \varphi\, dx = 1$ we can determine A, if it exists, in terms of the four parameters $a, c_2, \alpha_1, \alpha_2$; denote it therefore by A $(a \cdot c_2 \cdot \alpha_1 \cdot \alpha_2)$

Comparing with Koopman's form, it is evident that neither α_1 nor α_2 admits of estimation by sufficient statistics. If, however, the population values of these parameters are known, we may estimate the other two by sufficient statistics; thus, since

$$\varphi = \exp\left\{\frac{1}{c_2}\log\left[\frac{(x-\alpha_2)^{\alpha_2}}{(x-\alpha_1)^{\alpha_1}}\right]^{\frac{1}{\alpha_1-\alpha_2}} - \frac{a}{c_2}\log\left(\frac{x-\alpha_2}{x-\alpha_1}\right)^{\frac{1}{\alpha_1-\alpha_2}} + A(a\cdot c_2\cdot\alpha_1\cdot\alpha_2)\right\}$$

(i) $\frac{1}{n}\sum_{x} \log\left\{\frac{(x-\alpha_2)^{\alpha_2}}{(x-\alpha_1)^{\alpha_1}}\right\}^{\frac{1}{\alpha_1-\alpha_2}}$ is an unbiassed estimate of $-\frac{\partial A}{\partial(1/c_2)}$

(ii) $\frac{1}{n}\sum_{x} \log\left(\frac{x-\alpha_2}{x-\alpha_1}\right)^{\frac{1}{\alpha_1-\alpha_2}}$ is an unbiassed estimate of $+\frac{\partial A}{\partial(a/c_2)}$

in samples of n, and these statistics have minimum generalised variance.

(B) Denominator Possessing Equal Roots, say α. Then

$$\log \varphi = -\frac{1}{c_2}\int \frac{(x-a)\,dx}{(x-\alpha)^2}$$

$$\therefore \varphi = \exp\left\{-\frac{\alpha-a}{c_2}\cdot\frac{1}{x-\alpha} - \frac{1}{c_2}\log(x-\alpha) + A(a\cdot c_2\cdot\alpha)\right\}$$

where A, the constant of integration, is determined in terms of a, c_2, α by the total probability condition. By inspection, no sufficient statistic exists for the estimation of α.

When the population value of α is known

(i) $\frac{1}{n}\sum_{x}\frac{1}{x-\alpha}$ is an unbiassed and sufficient estimate of $-\partial A/\partial(a/c_2)$.

(ii) $\frac{1}{n}\left[\sum_{x}\left\{\log \overline{x-\alpha} + \frac{\alpha}{x-\alpha}\right\}\right]$ is an unbiassed and sufficient estimate of $\partial A/\partial(1/c_2)$.

(C) Constant $c_2 = 0$: Suppose first that $c_1/c_0 \neq -1/a$.

We have

$$\log \varphi = -\frac{1}{c_1}\int\frac{(x-a)\,dx}{x-b} \qquad \text{where } b = -\frac{c_0}{c_1}.$$

Hence

$$\varphi = \exp\left\{-\frac{x}{c_1} + \frac{b-a}{c_1}\log(x_1-b) + A(a\cdot b\cdot c_1)\right\}$$

(A has the usual significance).

No sufficient statistic exists for estimating b.

When the population value of b is known, two sufficient statistics can be obtained, e.g.,

$\frac{1}{n}\sum_x \log(x-b)$; an unbiassed estimate of $-\partial A/\partial\left(\frac{b-a}{c_1}\right)$.

$\frac{1}{n}\sum x$; an unbiassed estimate of $\partial A/\partial(1/c_1)$.

In this case, the mean emerges as sufficient. If the range is (b,∞) - where b is known - the total probability condition demands

$$\int_b^\infty (x-b)^{\frac{b-a}{c_1}} e^{-x/c_1}\,dx = e^{-A}$$

whence, integrating

$$A = \frac{b}{c_1} - \left(\frac{b-a}{c_1}+1\right)\log c_1 - \log\Gamma\left(\frac{b-a}{c_1}+1\right)$$

Consequently (choosing $1/c_1$ and $\overline{b-a}/c_1$ as independent variables ψ_1, ψ_2) the mean estimates $b + \left(\frac{b-a}{c_1}+1\right)c_1 = 2b + c_1 - a$.

If, in the original distribution, $c_2 = 0,\ c_1 \neq 0$ but $\frac{c_1}{c_0} = -\frac{1}{a}$ we have

$$\frac{d\varphi}{dx} = -\frac{\varphi_1}{c_1} \qquad \text{or} \quad \varphi = \text{constant},$$

a case of no interest.

(D) <u>Constants $c_2 = c_1 = 0$;</u> Here

$$\log\varphi = -\int\frac{(x-a)\,dx}{c_0}$$

$$\therefore\ \varphi = A e^{-(x-a)^2/2c_0}$$

If the range of x is $\pm\infty$ the constant $A = (2\pi c_0)^{-\frac{1}{2}}$ and the distribution is normal. As we have seen, two sufficient statistics exist; moreover, the mean is a sufficient estimate of a, which 'locates' the curve.

(E) <u>Denominator Possessing Complex Roots</u>:- We have

$$\log \varphi = -\frac{1}{c_2}\int \frac{(x-a)\,dx}{\left(x+\frac{c_1}{2c_2}\right)^2 + \frac{c_0}{c_2} - \frac{c_1^2}{4c_2^2}}$$

whence

$$\varphi = \exp\left[-\frac{1}{2c_2}\log\left\{\overline{x+\frac{c_1}{2c_2}}^{\,2} + \frac{4c_0c_2 - c_1^2}{4c_2^2}\right\} + \frac{2a}{(4c_0c_2-c_1^2)^{1/2}}\tan^{-1}\frac{2c_2x+c_1}{(4c_0c_2-c_1^2)^{1/2}} + A(a\cdot c_0\cdot c_1\cdot c_2)\right]$$

No sufficient statistics exist for the estimation of any of c_0, c_1, c_2. When the population values of the c's are given, there is a sufficient statistic for the estimation of a. In particular,

$\frac{1}{n}\sum_x \tan^{-1}\frac{2c_2x+c_1}{(4c_0c_2-c_1^2)^{1/2}}$ is an unbiassed estimate of $-\frac{(4c_0c_2-c_1^2)^{1/2}}{2}\cdot\frac{\partial A}{\partial a}$

and has minimum variance.

To sum up - the Pearson curves admit sufficient statistics for only two of the four parameters (if the denominator of $d\varphi/dx$ has real roots) or for only one parameter (if the denominator has complex roots) - the other two, or three, coefficients being regarded as known. For only two distributions (normal and Gamma) is the mean sufficient; and in only the former case does it estimate the parameter of location.

5.5 Koopman's Two-Parameter Distribution and Maximum Likelihood:- In applying the method of Maximum Likelihood to the distribution

$$\varphi = \exp\{\psi_1 f_1(x) + \psi_2 f_2(x) + F(\psi_1\cdot\psi_2) + c(x)\}$$

we must first decide what coefficients we wish to estimate.

Let these be $\theta_1(\psi_1 \cdot \psi_2)$ and $\theta_2(\psi_1 \cdot \psi_2)$. The likelihood function, for a Bernouillian sample of n, is

$$L = \psi_1 \sum_i f_1(x_i) + \psi_2 \sum_i f_2(x_i) + nF(\psi_1 \cdot \psi_2) + \sum_i c(x_i)$$

The precept $\frac{\partial L}{\partial \theta_1} = \frac{\partial L}{\partial \theta_2} = 0$ therefore yields

$$\frac{\partial \psi_1}{\partial \theta_j} \sum_i f_1(x_i) + \frac{\partial \psi_2}{\partial \theta_j} \sum_i f_2(x_i) + n \frac{\partial F}{\partial \theta_j} = 0 \quad [j = 1, 2]$$

or, since $\frac{\partial F}{\partial \theta_j} = \frac{\partial F}{\partial \psi_1} \frac{\partial \psi_1}{\partial \theta_j} + \frac{\partial F}{\partial \psi_2} \frac{\partial \psi_2}{\partial \theta_j}$,

$$\left.\begin{aligned} \left(\sum_i f_1(x_i) + n\frac{\partial F}{\partial \psi_1}\right)\frac{\partial \psi_1}{\partial \theta_1} + \left(\sum_i f_2(x_i) + n\frac{\partial F}{\partial \psi_2}\right)\frac{\partial \psi_2}{\partial \theta_1} = 0 \\ \left(\sum_i f_1(x_i) + n\frac{\partial F}{\partial \psi_1}\right)\frac{\partial \psi_1}{\partial \theta_2} + \left(\sum_i f_2(x_i) + n\frac{\partial F}{\partial \psi_2}\right)\frac{\partial \psi_2}{\partial \theta_2} = 0 \end{aligned}\right\}$$

Hence the maximum likelihood equations are

$$\sum_i f_1(x_i) + n\frac{\partial F}{\partial \psi_1} = 0 \;;\; \sum_i f_2(x_i) + n\frac{\partial F}{\partial \psi_2} = 0$$

which are independent of the parameters θ_1, θ_2. If we choose $\theta_1 = -\partial F/\partial \psi_1$, $\theta_2 = -\partial F/\partial \psi_2$ we obtain precisely the same statistics as in Section 5.3.2. Thus Maximum Likelihood can be made - by suitably selecting the coefficients to be estimated - to yield unbiassed statistics of minimum generalised variance. This result is not surprising, in view of the theorem discovered by Geary ("The Estimation of Many Parameters," Journal of The Royal Statistical Society, Vol. 55, Part 3, 1942), "in indefinitely large random samples, the maximum likelihood estimates of the parameters give minimum generalised variance."

5.5.1 Unbiassed Maximum Likelihood Statistics:-

θ_1, θ_2 could have been chosen as functions other than $-\partial F/\partial\psi_1$, $-\partial F/\partial\psi_2$ respectively. If θ_1 were, say, a given function $\mathcal{G}(-\partial F/\partial\psi_1, -\partial F/\partial\psi_2)$ we would have to solve the foregoing maximum likelihood equations accordingly, whereupon we would obtain $\mathcal{G}\left(\frac{1}{n}\sum_i f_1(x_i), \frac{1}{n}\sum_i f_2(x_i)\right)$ as the required statistic.

Let us determine what functions G can be estimated by maximum likelihood statistics which are unbiassed. Writing T_j for $\frac{1}{n}\sum_i f_j(x_i)$ $[j = 1, 2]$ we have

$$\frac{\partial L}{\partial\psi_j} = nT_j + n\frac{\partial F}{\partial\psi_j}$$

Therefore

$$\mathcal{G}(T_1, T_2) = \mathcal{G}\left(\overline{-\frac{\partial F}{\partial\psi_1} + \frac{1}{n}\frac{\partial L}{\partial\psi_1}}, \overline{-\frac{\partial F}{\partial\psi_2} + \frac{1}{n}\frac{\partial L}{\partial\psi_2}}\right)$$

When G possesses partial derivatives up to the third order

$$\begin{aligned}\mathcal{G}(T_1 \cdot T_2) &= \mathcal{G}\left(-\frac{\partial F}{\partial\psi_1}, \frac{\partial F}{\partial\psi_2}\right) + \frac{1}{n}\left(\frac{\partial L}{\partial\psi_1}\frac{\partial\hat{\mathcal{G}}}{\partial T_1} + \frac{\partial L}{\partial\psi_2}\frac{\partial\hat{\mathcal{G}}}{\partial T_2}\right)\\ &+ \frac{1}{2n^2}\left[\left(\frac{\partial L}{\partial\psi_1}\right)^2\frac{\partial^2\hat{\mathcal{G}}}{\partial T_1^2} + 2\frac{\partial L}{\partial\psi_1}\cdot\frac{\partial L}{\partial\psi_2}\frac{\partial^2\hat{\mathcal{G}}}{\partial T_1\partial T_2} + \left(\frac{\partial L}{\partial\psi_2}\right)^2\frac{\partial^2\hat{\mathcal{G}}}{\partial T_2^2}\right]\\ &+ \frac{1}{3!\,n^3}\left[\left(\frac{\partial L}{\partial\psi_1}\right)^3\frac{\partial^3}{\partial T_1^3} + 3\left(\frac{\partial L}{\partial\psi_1}\right)^2\frac{\partial L}{\partial\psi_2}\frac{\partial^3}{\partial T_1^2\partial T_2} + 3\frac{\partial L}{\partial\psi_1}\left(\frac{\partial L}{\partial\psi_2}\right)^2\frac{\partial^3}{\partial T_1\partial T_2^2}\right.\\ &\left.+ \left(\frac{\partial L}{\partial\psi_2}\right)^3\frac{\partial^3}{\partial T_2^3}\right]\mathcal{G}\left(-\frac{\partial F}{\partial\psi_1} + \frac{\theta}{n}\frac{\partial L}{\partial\psi_1}, -\frac{\partial F}{\partial\psi_2} + \frac{\theta'}{n}\frac{\partial L}{\partial\psi_2}\right) \qquad (25)\end{aligned}$$

where $0 \leq \theta < 1$; $0 \leq \theta' < 1$. The circumflex denotes that, after differentiation, the arguments are given the values $T_1 = -\partial F/\partial\psi_1$, $T_2 = -\partial F/\partial\psi_2$.

The expectation, over all possible samples of n, of the maximum likelihood estimate $G(T_1 \cdot T_2)$ is

$$E'(\mathcal{G}) = \int \mathcal{G}(T_1 \cdot T_2)\, \Phi\, dx' \quad .$$

Now, differentiating the equation

$$\int \Phi\, dx' = 1$$

with respect to ψ_i , we have

$$\int \left(\sum_j f_i(x_j) + n \frac{\partial F}{\partial \psi_i} \right) \Phi\, dx' = 0$$

$$\text{or} \quad \int \frac{\partial L}{\partial \psi_i}\, \Phi\, dx' = 0 \qquad [i = 1, 2] \tag{26}$$

Differentiating again with respect to ψ_j gives

$$\int \frac{\partial L}{\partial \psi_i} \frac{\partial L}{\partial \psi_j}\, \Phi\, dx' + \int \frac{\partial^2 L}{\partial \psi_i \partial \psi_j}\, \Phi\, dx' = 0$$

$$\text{or} \quad \int \frac{\partial L}{\partial \psi_i} \frac{\partial L}{\partial \psi_j}\, \Phi\, dx' = -n \frac{\partial^2 F}{\partial \psi_i \partial \psi_j} \quad [i, j = 1, 2] \tag{27}$$

Multiply (25) by Φ, integrate, and simplify by means of the relations just derived. We find

$$\begin{aligned} E'(\mathcal{G}) = {} & \mathcal{G}(-\partial F/\partial \psi_1 \, , \, -\partial F/\partial \psi_2) \\ & - \frac{1}{2n} \left[\frac{\partial^2 F}{\partial \psi_1^2} \frac{\partial^2 \hat{\mathcal{G}}}{\partial T_1^2} + 2 \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} \frac{\partial^2 \hat{\mathcal{G}}}{\partial T_1 \partial T_2} + \frac{\partial^2 F}{\partial \psi_2^2} \frac{\partial^2 \hat{\mathcal{G}}}{\partial T_2^2} \right] \\ & + O(n^{-2}) \end{aligned} \tag{28}$$

Since $\mathcal{G}(T_1 \cdot T_2)$ is an estimate of $\mathcal{G}\left(-\frac{\partial F}{\partial \psi_1}, -\frac{\partial F}{\partial \psi_2}\right)$ this statistic is biassed unless the second and subsequent terms are zero. The necessary and sufficient condition for this is clearly

$$\frac{\partial^2 \hat{\mathcal{G}}}{\partial T_1^2} = \frac{\partial^2 \hat{\mathcal{G}}}{\partial T_1 \partial T_2} = \frac{\partial^2 \hat{\mathcal{G}}}{\partial T_2^2} = 0$$

the most general solution of which is

$$g(T_1, T_2) = AT_1 + BT_2 + C.$$

(A, B, C constants). Thus the only unbiassed statistics yielded by maximum likelihood for the Koopman distribution are

$$\frac{1}{n}\sum_i f_1(x_i) \quad , \quad \frac{1}{n}\sum_i f_2(x_i)$$

and linear combinations thereof.

Equation (28) indicates that in large samples

$$E'(g) - g\left(-\frac{\partial F}{\partial \psi_1}, -\frac{\partial F}{\partial \psi_2}\right) = O(n^{-1})$$

$$\rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

This proves that any maximum likelihood statistic for the Koopman distribution is <u>consistent</u>.

<u>5.5.2</u> <u>Variances and Covariance by Likelihood Theory</u>:- The general theory of Maximum Likelihood gives the following precept for calculating the variances and covariance, in large samples, of the estimates of θ_1, θ_2: calculate the expectation, over all samples of n, of $\partial^2 L/\partial\theta_i\partial\theta_j$, denoting it by $\overline{\partial^2 L/\partial\theta_i\partial\theta_j}$ $[i, j = 1, 2]$ Form the Hessian determinant

$$H = \left| \overline{\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}} \right|$$

Then the variance of the estimate of θ_i, in large samples, is

$$\frac{1}{H}\left(\text{cofactor of } \overline{\partial^2 L / \partial \theta_i^2}\right) \qquad [i = 1, 2]$$

and the covariance of the two estimates, in large samples, is

$$\frac{1}{H}\left(\text{cofactor of } \overline{\partial^2 L / \partial \theta_1 \partial \theta_2}\right) .$$

We apply this rule for the choices

$$\theta_i = -\frac{\partial F}{\partial \psi_i} .$$

From the likelihood function, we derive

$$\frac{\partial L}{\partial \psi_i} = \frac{\partial L}{\partial \theta_1}\frac{\partial \theta_1}{\partial \psi_i} + \frac{\partial L}{\partial \theta_2}\frac{\partial \theta_2}{\partial \psi_i}$$

$$\doteq -\frac{\partial^2 F}{\partial \psi_1 \partial \psi_i}\frac{\partial L}{\partial \theta_1} - \frac{\partial^2 F}{\partial \psi_2 \partial \psi_i}\frac{\partial L}{\partial \theta_2} \qquad [i = 1, 2]$$

$$\frac{\partial^2 L}{\partial \psi_i \partial \psi_j} \doteq -\frac{\partial^3 F}{\partial \psi_1 \partial \psi_i \partial \psi_j}\frac{\partial L}{\partial \theta_1} - \frac{\partial^3 F}{\partial \psi_2 \partial \psi_i \partial \psi_j}\frac{\partial L}{\partial \theta_2}$$

$$+ \frac{\partial^2 F}{\partial \psi_1 \partial \psi_i}\left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_j}\frac{\partial^2 L}{\partial \theta_1^2} + \frac{\partial^2 F}{\partial \psi_2 \partial \psi_j}\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}\right)$$

$$+ \frac{\partial^2 F}{\partial \psi_2 \partial \psi_i}\left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_j}\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2} + \frac{\partial^2 F}{\partial \psi_2 \partial \psi_j}\frac{\partial^2 L}{\partial \theta_2^2}\right)$$

$$[i, j = 1, 2]$$

Multiply $\frac{\partial L}{\partial \psi_1}$, $\frac{\partial L}{\partial \psi_2}$ by Φ and integrate.

By (26), there results

$$\overline{\frac{\partial L}{\partial \theta_1}} = 0$$

$$\overline{\frac{\partial L}{\partial \theta_2}} = 0$$

Multiply each of $\partial^2 L/\partial\psi_1^2$, $\partial^2 L/\partial\psi_1\partial\psi_2$, $\partial^2 L/\partial\psi_2^2$ by Φ and integrate. We obtain

$$n\frac{\partial^2 F}{\partial \psi_1^2} = \left(\frac{\partial^2 F}{\partial \psi_1^2}\right)^2 \overline{\frac{\partial^2 L}{\partial \theta_1^2}} + 2\frac{\partial^2 F}{\partial \psi_1^2}\,\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\,\overline{\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}} + \left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\right)^2 \overline{\frac{\partial^2 L}{\partial \theta_2^2}}$$

$$n\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} = \frac{\partial^2 F}{\partial \psi_1^2}\,\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\,\overline{\frac{\partial^2 L}{\partial \theta_1^2}} + \left\{\frac{\partial^2 F}{\partial \psi_1^2}\,\frac{\partial^2 F}{\partial \psi_2^2} + \left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\right)^2\right\}\overline{\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}} + \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\,\frac{\partial^2 F}{\partial \psi_2^2}\,\overline{\frac{\partial^2 L}{\partial \theta_2^2}}$$

$$n\frac{\partial^2 F}{\partial \psi_2^2} = \left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\right)^2 \overline{\frac{\partial^2 L}{\partial \theta_1^2}} + 2\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}\,\frac{\partial^2 F}{\partial \psi_2^2}\,\overline{\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}} + \left(\frac{\partial^2 F}{\partial \psi_2^2}\right)^2 \overline{\frac{\partial^2 L}{\partial \theta_2^2}}$$

Solving this system of linear equations,

$$\overline{\frac{\partial^2 L}{\partial \theta_1^2}} = n \frac{\partial^2 F}{\partial \psi_2^2} \Big/ \Delta$$

$$\overline{\frac{\partial^2 L}{\partial \theta_1 \partial \theta_2}} = -n \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} \Big/ \Delta$$

$$\overline{\frac{\partial^2 L}{\partial \theta_2^2}} = n \frac{\partial^2 F}{\partial \psi_1^2} \Big/ \Delta$$

$$\Delta = \frac{\partial^2 F}{\partial \psi_1^2} \cdot \frac{\partial^2 F}{\partial \psi_2^2} - \left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} \right)^2$$

The Hessian determinant can now be written down:

$$H = \begin{vmatrix} \overline{\dfrac{\partial^2 L}{\partial \theta_1^2}} & \overline{\dfrac{\partial^2 L}{\partial \theta_1 \partial \theta_2}} \\ \overline{\dfrac{\partial^2 L}{\partial \theta_2 \partial \theta_1}} & \overline{\dfrac{\partial^2 L}{\partial \theta_2^2}} \end{vmatrix}$$

$$= \frac{n^2}{\Delta^2} \left\{ \frac{\partial^2 F}{\partial \psi_1^2} \cdot \frac{\partial^2 F}{\partial \psi_2^2} - \left(\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} \right)^2 \right\}$$

$$= \frac{n^2}{\Delta}$$

similarity - as well as an instructive analogy with the theory of linear estimation by Least Squares - has been discussed elsewhere ("On the Estimation of Statistical Parameters," A. C. Aitken and H. Silverstone, Proceedings of the Royal Society of Edinburgh, Section A, Vol. 61). Despite the final agreement, however, it msy be worth while to dwell for a moment on some points of difference, both in method and in conclusions.

5.6.1 Two important features of estimation by unbiassed statistics of minimum generalised variance concern (i) the distribution of the estimates and (ii) the treatment of the size of sample. As regards (i), no assumption whatsoever was made. The distribution of the estimate, be it normal or not, symmetrical or skew, is immaterial to the development of our theory. In Maximum Likelihood, on the other hand, the requirement of normally distributed statistics is fundamental. The precept for obtaining variances and covariances, for instance, is inextricably bound up with this assumption. (See, e.g., R. A. Fisher's paper, "The Logic of Inductive Inference," Journal of the Royal Statistical Society, Vol. 98, wherein he says - p.42 - "if T be an estimate of an unknown parameter θ, normally distributed with variance V, then the limit of $\frac{1}{nV}$ ------ ")

Because of the necessity for having normally distributed statistics, Maximum Likelihood distinguishes between large and small samples, and many of its theorems are valid only for the former. To quote Fisher again (loc. cit. p.41) "this.....naturally requires that our edifice shall be built in two stories. In the first we are concerned with the theory of large samples..... In the second story, where the real problem of finite samples is considered....." Estimation by unbiassed statistics of minimum generalised variance, in contrast, necessitates no qualification regarding sample size. This quantity, we recall, only affects the analysis by fixing the number of integrations to be performed.

<u>5.6.2</u> By handling expressions which are true for any size of sample, we detected (Section 5.1.1) the presence of a weak minimum. By following the usual Maximum Likelihood technique, and studying only "indefinitely large random samples," Geary was unable to state whether the resulting statistics made the generalised variance a strong or a weak minimum. He said merely that it was a "minimum" - unqualified - as, of course, is correct for the limiting case considered.

<u>5.6.3</u> The phenomenon of basic parametric coefficients - which emerges clearly from the methods of estimation here developed - is not brought out either by Maximum Likelihood, or by the criterion of sufficiency; yet it is an aspect of the theory which demands consideration, and which may modify previously accepted views - even, as we noted in Section 5.2, on such a well-worn subject as the normal curve of error.

The virtue, in this connection, of our two postulates is that they automatically reveal the coefficient to be estimated, without the introduction of any additional criterion.

5.6.4 The proponent of Maximum Likelihood might point out that unbiassed statistics of minimum generalised variance only exist for certain distributions - Koopman's type - whereas his method produces a result in many more cases. This is true, and a legitimate criticism. It would seem that the advantages adumbrated above are obtained at the expense of a limitation of applicability. From one point of view, indeed, our criteria determine not so much statistics, as distributions admitting certain species of estimates. (See Section 5.3, in which the equations (14) were treated as differential equations obeyed by "acceptable" probability distributions). This does however show - a partial reply to the foregoing criticism - that for non-Koopman forms there are no statistics possessing the desired attributes. Maximum Likelihood always yields sufficient statistics when they exist (Koopman's Third Theorem) but does not in itself tell whether a particular estimate is sufficient or not.

5.6.5 Another criticism which can be levelled against the present method is the restriction imposed on the arbitrary functions $f_1(x'), f_2(x')$ of Section 5.1. The restriction is however formal rather than real. When the expression μ_r (Section 5.1) is evaluated for Koopman's distribution, it is found to be a polynomial in $T_1 = \frac{1}{n}\sum_i f_1(x_i)$ and $T_2 = \frac{1}{n}\sum_i f_2(x_i)$, i.e., one part of f is any arbitrary function of T_1, T_2 capable of expansion in a Maclaurin series. The other part, involving as it does the solutions of a certain integral equation, represents a double infinity of functions. In any event, restrictions of this nature are not new in statistical analysis. Maximum Likelihood theory itself, for instance, contains such a feature. Thus Fisher (loc. cit. p.45) proves that "of the methods of estimation based on linear functions of the frequencies, that with smallest limiting variance is the method of maximum likelihood." Similarly, in his paper we have referred to, Geary compares the generalised variance of maximum likelihood statistics with that of others based on linear functions of the frequencies.

5.6.6 A final point. In Maximum Likelihood the emphasis rests on estimating a coefficient by solving an equation. One arrives at a relation $w(x'|\hat{\theta})=0$ say, which must be manipulated by ordinary algebraic or analytic technique, so that $\hat{\theta}$ is obtained in terms of the observations x'. In distinction, the method of unbiassed minimum generalised variance leads, not to an equation, but to a functional form. If a certain operation gives a function of x' alone, we have the desired statistic; if it does not yield such a form, the statistic does not exist.

CHAPTER SIX

THE SAMPLING DISTRIBUTION OF CERTAIN STATISTICS

6.0 While a formal distinction is drawn between the problems of distribution and of estimation, no clear boundary can be traced in practice, and the solution of the former may be a necessary preliminary to the study of the latter. For instance, the test of the significance of an estimate - or more commonly of the difference of two estimates - requires a knowledge of the distribution of the statistic in question.

6.1 The Sampling Distribution of an Unbiassed Statistic of Minimum Variance:- The one-parameter distribution

$$\varphi = \exp\{\psi f(x) + F(\psi) + c(x)\}$$

admits the statistic $T = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$ as an unbiassed estimate of $-dF/d\psi$, possessing minimum variance.

The moment generating function (M.G.F.) of T in Bernouillian samples of n is

$$M(\alpha) = \int\cdots\int e^{\alpha T}\,\Phi\,dx'$$

$$= \int\cdots\int \exp\left\{\frac{\alpha}{n}\sum f(x) + \psi\sum f(x) + nF(\psi) + \sum c(x)\right\}dx'$$

$$= \left[\int \exp\left\{\left(\psi + \frac{\alpha}{n}\right)f(x) + F(\psi) + c(x)\right\}dx\right]^n$$

By the total probability condition

$$\int \exp\{\psi f(x) + F(\psi) + c(x)\}\, dx = 1$$

for all values of ψ. This relation consequently holds when ψ is replaced by $\psi + \frac{\alpha}{n}$, whence

$$M(\alpha) = \left[\exp\{F(\psi) - F(\psi+\tfrac{\alpha}{n})\}\int \exp\{(\psi+\tfrac{\alpha}{n}) f(x) + F(\psi+\tfrac{\alpha}{n}) + c(x)\}\, dx\right]^n$$

$$= \exp\left\{ n\,F(\psi) - nF(\psi + \tfrac{\alpha}{n}) \right\} \quad . \qquad (1)$$

The seminvariant generating function (S.G.F.) of T is

$$L(\alpha) = \log M(\alpha) = nF(\psi) - n\,F(\psi + \tfrac{\alpha}{n}) . \qquad (2)$$

<u>6.1.1</u> <u>Moments and Seminvariants of T</u> :- The r^{th} moment about the origin of T, say μ'_r, is the coefficient of $\alpha^r/r!$ in the M.G.F. That is,

$$\mu'_r = \left[\frac{d^r M(\alpha)}{d\alpha^r}\right]_{\alpha=0}$$

on expanding M as a Maclaurin Series in powers of α.

By (1), $$\mu'_r = e^{nF(\psi)}\left[\frac{d^r e^{-nF(\psi+\frac{\alpha}{n})}}{d\alpha^r}\right]_{\alpha=0}$$

$$= \frac{1}{n^r} e^{nF(\psi)} \frac{d^r e^{-nF(\psi)}}{d\psi^r} \qquad (3)$$

Similarly, the r^{th} seminvariant of T, say λ_r, is the coefficient of $\alpha^r/r!$ in the S.G.F. Provided F is capable of expansion in a Maclaurin series (terminating or infinite), we find from (2) that

$$\lambda_r = -\frac{1}{n^{r-1}} \frac{d^r F(\psi)}{d\psi^r} \qquad (4)$$

The first two seminvariants of T are its mean and variance respectively. Formula (4), for $r=1$ and $r=2$ is therefore in agreement with the results of Chapter Three. The simple statistical interpretation of the successive derivatives of F, which (4) provides, is noteworthy.

Example:- The normal distribution with unit variance is

$$\varphi = \frac{1}{\sqrt{2\pi}} \exp\left\{ - \frac{(x-\psi)^2}{2} \right\}$$

$$= \exp\left\{ \psi x - \frac{\psi^2}{2} - \frac{x^2}{2} - \frac{1}{2}\log 2\pi \right\} .$$

By choosing $n=1$, $F(\psi) = -\frac{\psi^2}{2}$ in formula (3), we obtain the moments of x about the origin. The r^{th} moment is

$$\mu'_r = e^{-\frac{\psi^2}{2}} \left[d^r e^{\psi^2/2} / d\psi^r \right]$$

or, writing $\psi^2/2 = -t^2$,

$$\mu'_r = (-i\sqrt{2})^r e^{t^2} \left(d^r e^{-t^2} / dt^r \right) = (-i\sqrt{2})^r H_r(t)$$

$$= (-i\sqrt{2})^r H_r\left(i\psi/\sqrt{2} \right) ,$$

where H_r denotes Hermite's Polynomial of degree r.

6.1.2 The Sampling Distribution of T :- Let the sampling distribution of $T = \frac{1}{n}\sum_i f(x_i)$ be $g(T)$. The M.G.F. is accordingly

$$\int e^{\alpha T} g(T)\, dT$$

and, comparing with (1), we have the integral equation

$$\int e^{\alpha T} g(T)\, dT = \exp\left\{ nF(\psi) - nF\left(\psi + \frac{\alpha}{n}\right) \right\} \qquad (5)$$

holding for all values of α.

Suppose the limits of T are $\pm\infty$. Replacing α by $-\alpha$, (5) becomes

$$\int_{-\infty}^{\infty} e^{-\alpha T} g(T)\,dT = \exp\left\{nF(\psi) - nF\left(\psi - \frac{\alpha}{n}\right)\right\}$$

whence, inverting by Mellin's theorem,

$$g(T) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \exp\left\{zT + nF(\psi) - nF\left(\psi - \frac{z}{n}\right)\right\} \qquad (6)$$

where c is any real number which exceeds the real part of any pole of the integrand. By the general theory of transforms, $g(T)$ is unique.

When the range of T is finite, we must solve (5) by the general methods of Chapter Two. The function $g(T)$ so obtained is the only possible function of class L^2; in particular, if a continuous solution exists, it is the unique continuous solution.

Let us, in (5), put $\alpha = 0$. There results

$$\int g(T)\,dT = 1 .$$

Therefore $g(T)$ satisfies the total probability condition - as, of course, it must do.

<u>6.1.3</u> <u>Non-Normality of the Distribution of T :-</u> The M.G.F. of a normal variate T whose mean is m and whose variance is σ^2 is

$$M_1(\alpha) = \int_{-\infty}^{\infty} e^{\alpha T} \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-(T-m)^2/2\sigma^2}\,dT$$

$$= \exp\left(\frac{\alpha^2\sigma^2}{2} + \alpha m\right)$$

Suppose that our statistic T has range $\pm\infty$. We know that its mean is $m = -\frac{dF}{d\psi}$, and that its variance is

$\sigma^2 = -\frac{1}{n}\,\frac{d^2F}{d\psi^2}$. From (1), the condition that T be a normal variate is thus

$$\exp\left\{nF(\psi) - nF\left(\psi+\frac{\alpha}{n}\right)\right\} = \exp\left(-\frac{\alpha^2}{2n}\,\frac{d^2F}{d\psi^2} - \alpha\,\frac{dF}{d\psi}\right)$$

Assume that F possesses derivates up to the third order, so that

$$F\left(\psi+\frac{\alpha}{n}\right) = F(\psi) + \frac{\alpha}{n}\frac{dF}{d\psi} + \frac{\alpha^2}{2n^2}\,\frac{d^2F}{d\psi^2} + \frac{\alpha^3}{3!\,n^3}\,\frac{d^3F\left(\psi+\frac{\alpha\theta}{n}\right)}{d\psi^3}$$

$[0 \leq \theta < 1]$. Substituting in the preceding equation, we see that the necessary and sufficient condition that T be normally distributed is $d^3F/d\psi^3 = 0$ for all values of ψ.

i.e., $$F = A + B\psi + C\psi^2$$

where A, B, C are arbitrary constants. Unless the F of Koopman's distribution is of this particular form, the statistic T is not a normal variate.

We observe that

$$\frac{M(\alpha)}{M_1(\alpha)} = \frac{\exp\left\{nF(\psi) - nF\left(\psi+\frac{\alpha}{n}\right)\right\}}{\exp\left(-\alpha\frac{dF}{d\psi} - \frac{\alpha^2}{2n}\,\frac{d^2F}{d\psi^2}\right)} = \exp\left\{-\frac{\alpha^3}{3!\,n^2}\,\frac{d^3F\left(\psi+\frac{\alpha\theta}{n}\right)}{d\psi^3}\right\}$$

$$\rightarrow 1 \quad \text{as} \quad n \rightarrow \infty.$$

i.e., in indefinitely large samples, the M.G.F. of T tends to the M.G.F. of the normal variate with the same mean and variance.

6.1.4 The Order of Magnitude of the Moments of T:- The result of the last paragraph may be put: in indefinitely large samples, the distribution of T tends to be normal. We now derive a more precise statement of this property, by comparing the magnitudes of the moments of T and those of a normal variate with the same mean and variance.

The M.G.F. about the mean of a normal variate with variance $-\frac{1}{n}\frac{d^2F}{d\psi^2}$ is (see Section 6.1.3)

$$M_1'(\alpha) = \exp\left(-\frac{\alpha^2}{2n}\frac{d^2F}{d\psi^2}\right)$$

The moments about the mean of this variate are accordingly

$$\left.\begin{aligned} \mu^{x}_{2r} &= (-1)^r \frac{(2r)!}{(r!)^2}\frac{1}{n^r\cdot 2^r}\left(\frac{d^2F}{d\psi^2}\right)^r \\ \mu^{x}_{2r+1} &= 0 \end{aligned}\right\} \qquad (7)$$

As $n \to \infty$, μ^{x}_{2r} is thus $O(n^{-r})$, while μ^{x}_{2r+1} is accurately zero.

The M.G.F. of T about its mean $-\frac{dF}{d\psi}$ is (see equation (1))

$$\begin{aligned} M'(\alpha) &= \exp\left\{nF(\psi) - nF\left(\psi+\frac{\alpha}{n}\right) + \alpha\frac{dF}{d\psi}\right\} \\ &= \left(\exp -\frac{\alpha^2}{2n}\frac{d^2F}{d\psi^2}\right)\left(\exp\left\{-\frac{\alpha^3}{3!n^2}\frac{d^3F}{d\psi^3} - \frac{\alpha^4}{4!n^3}\frac{d^4F}{d\psi^4}\cdots\right\}\right) \\ &= M_1'(\alpha)\exp\left\{-\frac{\alpha^3}{3!n^2}\frac{d^3F}{d\psi^3} - \frac{\alpha^4}{4!n^3}\frac{d^4F}{d\psi^4}\cdots\right\} \\ &= M_1'(\alpha)\left[1 - \left(\frac{\alpha^3}{3!n^2}\frac{d^3F}{d\psi^3} + \frac{\alpha^4}{4!n^3}\frac{d^4F}{d\psi^4}\cdots\right)\right. \\ &\qquad \left. + \frac{1}{2}\left(\frac{\alpha^3}{3!n^2}\frac{d^3F}{d\psi^3} + \frac{\alpha^4}{4!n^3}\frac{d^4F}{d\psi^4}\cdots\right)^2 - \cdots\right] \end{aligned} \qquad (8)$$

Consider the term in α^{2r+1} say, which consists of a sum of products involving α^{2p} from the factor $M_1'(\alpha)$ and $\alpha^{2(r-p)+1}$ from the second factor ($p = 0, 1, 2, \cdots \overline{r-1}$.)
The latter contains contributions

of order $O\{n^{-2(r-p)}\}$ from the first bracket(
of order $O\{n^{-2(r-p)+1}\}$ from the second bracket
$\vdots$
of order $O\{n^{-2(r-p)-1+t}\}$ from the t^{th} bracket,

where $3(t+1) > 2(r-p)+1 \geqslant 3t$.

The lowest order of magnitude (in indefinitely large samples) which arises from the product involving α^{2p} in $M_1'(\alpha)$ and $\alpha^{2(r-p)+1}$ in the second factor, is therefore - remembering (7) -

$$O\{n^{-2r+p-1+t}\}$$

or, putting $2(r-p)+1 = 3t+\omega$ ($\omega = 0, 1, 2$ as the case may be)

it is
$$O\{n^{-(4r+2+\omega-p)/3}\}.$$

The negative exponent is least when p is as large as possible, viz., $p = r-1$, and when $\omega = 0$. The term is then $O\{n^{-r-1}\}$. That is, the order of magnitude of μ_{2r+1}, the odd moment about the mean of T, is

$\mu_{2r+1} = O\{n^{-r-1}\}$ in indefinitely large samples.

We recall (7), which shows that the corresponding moment $\mu^{\times}_{2r+1}$ of the normal variate is accurately zero.

In like manner, we deduce that

$$\mu_{2r} = O(n^{-r}) \quad , \text{ in indefinitely large samples,}$$

which is of the <u>same</u> order of magnitude as μ^{x}_{2r}.

6.1.5 Skewness and Kurtosis of the Distribution of T:-

Since the odd moments of T are not in general zero, the distribution of this variate is unsymmetrical. The skewness is given by

$$\beta_1 = \mu_3^{\,2} / \mu_2^{\,3}$$

From equation (8), it is found that

$$\mu_3 = -\frac{1}{n^2}\frac{d^3F}{d\psi^3}$$

so that

$$\beta_1 = -\frac{1}{n}\left(\frac{d^3F}{d\psi^3}\right)^2 \cdot \left(\frac{d^2F}{d\psi^2}\right)^{-3} \qquad (9)$$

Since, in practice, $d^2F/d\psi^2 = -n\mu_2$ is negative, we see that the distribution of T is positively or negatively skew (i.e., $+\sqrt{\beta_1} \gtrless 0$) according as $\dfrac{d^3F}{d\psi^3} \lessgtr 0$.

The kurtosis is given by

$$\beta_2 = \mu_4 / \mu_2^{\,2}$$

From (8),

$$\mu_4 = \mu_4^{x} - \frac{1}{n^3}\frac{d^4F}{d\psi^4} = \frac{3}{n^2}\left(\frac{d^2F}{d\psi^2}\right)^2 - \frac{1}{n^3}\frac{d^4F}{d\psi^4}$$

whence

$$\beta_2 = 3 - \frac{1}{n}\frac{d^4F}{d\psi^4}\left(\frac{d^2F}{d\psi^2}\right)^{-2} \qquad (10)$$

The distribution of T is thus platykurtic or leptokurtic $(\beta_2 - 3 \gtrless 0)$ according as $\dfrac{d^4F}{d\psi^4} \lessgtr 0$.

We note from (9) and (10) that the skewness and excess $(\beta_2 - 3)$ are each $O(n^{-1})$ in indefinitely large samples.

6.1.6 The Sampling Distribution of T is of Koopman's Form:- Since x is by hypothesis distributed in a Koopman form, so is $f(x)$. This follows from Section 3.5.1, where it was proved that Koopman's form is invariant under a transformation of the variate. Thus if the distribution of $f(x)$ is denoted by g, and if the limits of f are a, b, we have

$$\int_a^b e^{\alpha f(x)} g(f)\, df = \text{M. G. F. of } f(x)$$
$$= \exp\{F(\psi) - F(\psi+\alpha)\}.$$

and we know that the solution g is a Koopman distribution.

Consider now the variate $T_1 = \sum_{i=1}^{n} f(x_i)$.
Let its sampling distribution be $g_1(T_1)$ $[na \le T_1 \le nb]$ so that we have

$$\int_{na}^{nb} e^{\alpha T_1} g(T_1)\, dT_1 = \int \cdots \int e^{\alpha \Sigma f + \psi \Sigma f + nF + \Sigma c}\, dx'$$
$$= \left\{ \int e^{(\psi+\alpha) f(x) + F(\psi) + c(x)}\, dx \right\}^n$$
$$= \exp\{nF(\psi) - nF(\psi+\alpha)\}.$$

or $$\int_a^b e^{\alpha(T_1 n)} g(T_1 n)\, d(T_1 n) = \exp\{nF(\psi) - nF(\psi+\alpha)\}$$

This is the same integral equation as that given above, with F replaced by nF. Therefore $g_1(T_1 n)$ is of the same form as g viz., of Koopman's form, which is, moreover, preserved by the transformation $T = T_1/n$. Hence the distribution of T admits a sufficient statistic.

6.2 The Sanpling Distributions of Two Unbiassed Statistics of Minimum Generalised Variance:- The two-parameter distribution

$$\varphi = \exp\{\psi_1 f_1(x) + \psi_2 f_2(x) + F(\psi_1\cdot\psi_2) + c(x)\}$$

admits the statistics $T_1 = \frac{1}{n}\sum_{i=1}^{n} f_1(x_i)$; $T_2 = \frac{1}{n}\sum_{i=1}^{n} f_2(x_i)$ as unbiassed estimates of $-\frac{\partial F}{\partial\psi_1}$, $-\frac{\partial F}{\partial\psi_2}$ respectively, and they give minimum generalised variance.

The bivariate M.G.F. of the simultaneous sampling distribution of T_1 and T_2 is

$$M(\alpha\cdot\beta) = \int\cdots\int e^{\alpha T_1 + \beta T_2}\,\Phi\, dx'$$

$$= \int\cdots\int \exp\Big[\frac{\alpha}{n}\sum f_1(x) + \frac{\beta}{n}\sum f_2(x) + \psi_1\sum f_1(x) + \psi_2\sum f_2(x) + nF(\psi_1\cdot\psi_2) + \sum c(x)\Big] dx'$$

$$= \Big\{\Big[\exp F(\psi_1\cdot\psi_2) - F(\psi_1+\tfrac{\alpha}{n}\cdot\psi_2+\tfrac{\beta}{n})\Big]\Big[\int \exp\{(\psi_1+\tfrac{\alpha}{n})f_1(x) + (\psi_2+\tfrac{\alpha}{n})f_2(x) + F(\psi_1+\tfrac{\alpha}{n}\cdot\psi_2+\tfrac{\beta}{n}) + c(x)\}\,dx\Big]^n$$

$$= \exp\{nF(\psi_1\cdot\psi_2) - nF(\psi_1+\tfrac{\alpha}{n}\cdot\psi_2+\tfrac{\beta}{n})\} \qquad (11)$$

by the total probability condition.

The bivariate S.G.F. is (12)

$$L(\alpha\cdot\beta) = \log M(\alpha\cdot\beta) = nF(\psi_1\cdot\psi_2) - nF(\psi_1+\tfrac{\alpha}{n}\cdot\psi_2+\tfrac{\beta}{n}).$$

Assuming that F can be expanded in a Maclaurin series, the $(r,s)^{th}$ seminvariant λ_{rs}, or the coefficient of $\alpha^r\beta^s/r!s!$ is

$$\lambda_{rs} = -\frac{1}{n^{r+s-1}}\cdot\frac{\partial^{r+s}F}{\partial\psi_1^r\,\partial\psi_2^s} \qquad (13)$$

Since λ_{10}, λ_{01} are the means of T_1 and T_2 respectively; since λ_{20} and λ_{02} are the respective variances; and since λ_{11} is their covariance, formula (13) is consistent with the results of Chapter Five.

6.2.1 Non-Normality of the Simultaneous Distribution of T_1 and T_2 :- We know the means, variances and covariance of T_1 and T_2 (equation (13). If the simultaneous distribution of these statistics were normal, the bivariate M.G.F. about the origin would be

$$M^{x}(\alpha\cdot\beta) = exp\left\{-\alpha\frac{\partial F}{\partial\psi_1} - \beta\frac{\partial F}{\partial\psi_2} - \frac{\alpha^2}{2n}\frac{\partial^2 F}{\partial\psi_1^2} - \frac{\alpha\beta}{n}\frac{\partial^2 F}{\partial\psi_1\partial\psi_2} - \frac{\beta^2}{2n}\frac{\partial^2 F}{\partial\psi_2^2}\right\}$$

(See Aitken, "Statistical Mathematics," p.85). The actual bivariate M.G.F. is formula (11). Expanding F in a Maclaurin series as far as the third partial derivatives, we note that the necessary and sufficient condition for $\dot{M}(\alpha\cdot\beta) \equiv M^{x}(\alpha\cdot\beta)$ is

$$\frac{\partial^3 F}{\partial\psi_1^3} = \frac{\partial^3 F}{\partial\psi_1^2\partial\psi_2} = \frac{\partial^3 F}{\partial\psi_1\partial\psi_2^2} = \frac{\partial^3 F}{\partial\psi_2^3} = 0$$

the most general solution of which is

$$F = A_0 + B_1\psi_1 + B_2\psi_2 + C_1\psi_1^2 + C_2\psi_2^2 + C_3\psi_1\psi_2 ,$$

where the A, B's, and C's are arbitrary constants. Consequently, unless the F of Koopman's distribution is of this special form, T_1 and T_2 are not normal variates.

In general

$$\frac{M(\alpha\cdot\beta)}{M^{x}(\alpha\cdot\beta)} = exp\left[-\frac{1}{n^2}\left(\alpha\frac{\partial}{\partial\psi_1} + \beta\frac{\partial}{\partial\psi_2}\right)^3 F\left(\psi_1 + \frac{\alpha\theta_1}{n}, \psi_2 + \beta\frac{\theta_2}{n}\right)\right]$$

(where $0 \leq \theta_1 < 1$; $0 \leq \theta_2 < 1$)

$$\therefore \frac{M(\alpha \cdot \beta)}{M^x(\alpha \cdot \beta)} = O\left(e^{-n^{-2}}\right) \quad \text{in indefinitely large samples}$$

$$\longrightarrow 1 \qquad \text{as } n \to \infty.$$

Therefore, the simultaneous distribution of T_1, T_2 tends to be normal in infinitely large samples.

<u>6.2.2</u> <u>Uncorrelated Statistics</u>:- The statistics T_1, T_2 would be uncorrelated if we had $\lambda_{11} = 0$

i.e., $$\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} = 0$$

or $$F = \mathcal{U}(\psi_1) + V(\psi_2)$$

($\mathcal{U}, V$ are arbitrary functions). As we have seen, no Koopman distribution of this form exists, so that the unbiassed statistics of minimum generalised variance are always correlated.

<u>6.2.3</u> <u>Statistically Independent Estimates</u>:- Let the S.G.F's. of T_1 alone, and of T_2 alone, be denoted by $L(\alpha \cdot 0)$ and $L(0 \cdot \beta)$ respectively. They are found by putting $\beta = 0$ in (12) in the first case, and $\alpha = 0$ in the second. By definition, T_1, T_2 are statistically independent if

$$L(\alpha \cdot \beta) = L(\alpha \cdot 0) + L(0 \cdot \beta)$$

i.e., if

$$F(\psi_1 \cdot \psi_2) - F(\psi_1 + \tfrac{\alpha}{n} \cdot \psi_2 + \tfrac{\beta}{n}) = F(\psi_1 \cdot \psi_2) - F(\psi_1 + \tfrac{\alpha}{n} \cdot \psi_2) + F(\psi_1 \cdot \psi_2) - F(\psi_1 \cdot \psi_2 + \tfrac{\beta}{n})$$

Expanding, with respect to the first variable, by the Mean Value Theorem, and dividing by α/n, this condition becomes

$$\frac{\partial}{\partial \psi_1}\left\{F(\psi_1 \cdot \psi_2) - F(\psi_1 \cdot \psi_2 + \tfrac{\beta}{n})\right\} = \frac{\alpha}{2n} \frac{\partial^2}{\partial \psi_1^2}\left\{F(\psi_1 + \tfrac{\alpha \theta_1}{n} \cdot \psi_2 + \tfrac{\beta}{n}) - F(\psi_1 + \tfrac{\alpha \theta_1}{n} \cdot \psi_2)\right\}$$

$[0 \leq \theta_1 < 1]$

That is, a function of β = a function of α and β.

Each side is therefore a constant, and this constant has the value zero (put, say $\beta = 0$) So

$$\frac{\partial}{\partial \psi_1}\left\{F(\psi_1 \cdot \psi_2) - F(\psi_1 \cdot \psi_2 + \tfrac{\beta}{n})\right\} = 0$$

or $\quad F(\psi_1 \cdot \psi_2) - F(\psi_1 \cdot \psi_2 + \frac{\beta}{n}) =$ function of ψ_2 and β only.

Similarly $F(\psi_1 \cdot \psi_2) - F(\psi_1 + \frac{\alpha}{n} \cdot \psi_2) =$ function of ψ_1 and α only.

It follows that the condition for statistical independence is

$$F(\psi_1 \cdot \psi_2) = U(\psi_1) + V(\psi_2)$$

which never subsists. It is interesting, though unimportant, that if T_1 and T_2 could ever be uncorrelated, they would also be statistically independent.

<u>6.2.4</u> <u>Special Parametric Relations</u>:- In certain cases, special functions of the parameter ψ_1 will be independent of ψ_2. Thus consider the distribution

$$\varphi = \exp\left\{\psi_1 f_1(x) + \psi_2 f_2(x) + u(\psi_1)\, v(\psi_2) + \kappa(x)\right\}.$$

The S.G.F. of $T_1 = \frac{1}{n}\sum_i f_1(x_i)$ alone is

$$L(\alpha \cdot 0) = n\, v(\psi_2)\left\{u(\psi_1) - u(\psi_1 + \tfrac{\alpha}{n})\right\}.$$

The r^{th} seminvariant of T_1 is consequently

$$\lambda_{r \cdot 0} = -\frac{1}{n^{r-1}}\, v(\psi_2)\, \frac{d^r u(\psi_1)}{d\psi_1^{\,r}}$$

Hence, for any pair of integers r, s,

$$\lambda_{r \cdot 0} / \lambda_{s \cdot 0} = n^{s-r}\left(\frac{d^r u}{d\psi_1^{\,r}}\right) \Big/ \left(\frac{d^s u}{d\psi_1^{\,s}}\right)$$

is independent of the coefficient ψ_2.

6.3 Examples:-

6.3.1 The Normal Distribution with Unit Variance

$$\varphi = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2}\right\}$$
$$= \exp\left\{-mx - \tfrac{1}{2}m^2 - \tfrac{1}{2}x^2 - \tfrac{1}{2}\log 2\pi\right\}$$

The M.G.F. of $T = \frac{1}{n}\sum x$ is (equation (1) with $F(m) = -\frac{1}{2}m^2$))

$$M(\alpha) = \exp\left(\alpha m + \alpha^2/2n\right)$$

which is the M.G.F. of a normal variate whose mean is m and whose variance is $1/n$. The sampling distribution of T is therefore

$$g(T) = \frac{1}{\sqrt{2\pi n}} \cdot \exp\left\{-\frac{(T-m)^2}{2n}\right\}$$

6.3.2 The Pearson Curve of Gamma Type:-

$$\varphi = \frac{1}{a'\Gamma(p+1)}\left(\frac{x}{a'}\right)^p \cdot e^{-x/a'} \qquad (x \geq 0;\ p > 0)$$

Writing $a' = 1/a$, we have the canonical form

$$\varphi = \exp\left\{-ax + p\log x + (p+1)\log a - \log\Gamma(p+1)\right\}$$

The M.G.F. of $T_2 = \frac{1}{n}\sum \log x$ i.e., of the geometric mean - is

$$M(0\cdot\beta) = \exp\left\{n(p+1)\log a - n\log\Gamma(p+1) - n\left(p+\tfrac{\beta}{n}+1\right)\log a + n\log\Gamma\left(p+\tfrac{\beta}{n}+1\right)\right\}$$
$$= a^{-\beta}\left\{\frac{\Gamma(p+1+\beta/n)}{\Gamma(p+1)}\right\}^n$$

The M.G.F. of the arithmetic mean $T_1 = -\frac{1}{n}\sum x$ is

$$M(\alpha\cdot 0) = \left(\frac{a}{a+\alpha/n}\right)^{n(p+1)}$$

Let $g(T_1)$ be the sampling distribution of T_1, which varies between 0 and $-\infty$. Then

$$\int_{-\infty}^{0} e^{\alpha T_1} g(T_1)\, dT_1 = \left\{a/\left(a + \tfrac{\alpha}{n}\right)\right\}^{n(p+1)}$$

Change T_1 to $-T_1$. Since the probability that $-T_1$ lies between $-T_1 \mp \frac{1}{2} dT_1$ is $g(T_1)\,dT_1$,

$$\int_{0}^{\infty} e^{-\alpha T_1} g(T_1)\, dT_1 = \left\{a/\left(a + \tfrac{\alpha}{n}\right)\right\}^{n(p+1)}.$$

By Mellin's theorem, we have formally

$$g(T_1)\,dT_1 = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{z T_1}\left(\frac{a}{a+z/n}\right)^{n(p+1)} dz \qquad (14)$$

We obtain $g(T_1)$ by the following process:

(i) by evaluating the contour integral in certain special circumstances.

(ii) by proving that the probability function so found always has the required M.G.F., not merely under the special conditions of (i).

(iii) by invoking the theorem that two functions possessing the same Laplace transform are identical, we assure ourselves that the function obtained in (i) is the unique distribution of T_1. (For this theorem, see, e.g., Titchmarsh, "Theory of Fourier Integrals," p.164).

Now, when $p > 0$, $|a + z/n|^{-n(p+1)} \to 0$ uniformly, as $|z| \to \infty$, with respect to $\arg z$. So, when $\frac{\pi}{2} \leq \arg z \leq \frac{3\pi}{2}$,

$$\left| z e^{z T_1}\left(\frac{a}{a+z/n}\right)^{n(p+1)} \right| \to 0 \text{ uniformly, as } |z| \to \infty,$$

with respect to $\arg z$. The integral (14) is therefore equal to the integral of the same function round a contour Γ, where Γ is formed by the line $c \pm i\infty$ and the infinite semi-circle in the third and fourth quadrants. The only singularity within Γ is at $z = -an$. (Note that, as T_1 is always negative, its mean value, which is $-(p+1)/a$, is negative; since p is taken > 0, a is > 0 and the singularity is inside the contour).

Write $z' = z + an$; the integrand of (14) becomes

$$e^{(z'-an)T_1}\left(\frac{an}{z'}\right)^{n(p+1)} = e^{-anT_1}(an/z')^{n(p+1)}\left\{1 + zT_1 + \cdots + \frac{z^k T_1^k}{k!}\cdots\right\}$$

Now for the special case $n(p+1)$ = integer, the point $z' = 0$ is a pole of order $n(p+1)$, where the residue is

$$e^{-anT_1}(an)^{n(p+1)}\frac{T_1^{np}}{(np)!}$$

Hence, from (14),

$$g(T_1) = \frac{(an)^{n(p+1)}}{(np)!}T_1^{np}e^{-anT_1} \qquad (np \text{ integral})$$

This suggests that, whatever the value of p, we should consider the probability distribution

$$g(T_1) = \frac{(an)^{n(p+1)}}{\Gamma(np+1)}T_1^{np}e^{-anT_1} \tag{15}$$

We have $\displaystyle\int_{-\infty}^{0} e^{\alpha T_1} g(T_1)\,dT_1 = \{an/(an+\alpha)\}^{n(p+1)}$ provided $p > 0$.

$$= M(\alpha \cdot 0)$$

Therefore, (15) represents the unique sampling distribution of the arithmetic mean in the Gamma frequency curve.

6.3.3 The Normal Curve of Error:-

$$\varphi = \frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}$$

Writing $\psi_1 = 1/2\sigma^2$, $\psi_2 = m/\sigma^2$ we obtain

$$\varphi = \exp\left\{-\psi_1 x^2 + \psi_2 x - \frac{\psi_2^2}{4\psi_1} + \frac{1}{2}\log\psi_1 - \frac{1}{2}\log 2\pi\right\}$$

which is of Koopman's form with

$$F(\psi_1\cdot\psi_2) = -\frac{\psi_2^2}{4\psi_1} + \frac{1}{2}\log\psi_1$$

The M.G.F. of $T_2 = \frac{1}{n}\sum x$ is thus

$$M(0\cdot\beta) = \exp\left\{\frac{\psi_2}{2\psi_1}\beta + \frac{\beta^2}{4n\psi_1}\right\}$$

which shows that T_2 is a normal variate with mean $\frac{\psi_2}{2\psi_1} = m$ and variance $\frac{1}{2n\psi_1} = \frac{\sigma^2}{n}$. This merely amplifies the result of Section 6.3.1.

The M.G.F. of $T_1 = -\frac{1}{n}\sum x^2$ is, by the usual formula,

$$M(\alpha\cdot 0) = \left(1+\frac{\alpha}{n\psi_1}\right)^{-n/2}\exp\left\{\frac{-\alpha\psi_2^2}{4\psi_1(\psi_1+\alpha/n)}\right\}$$

The range of T_1 is 0 to $-\infty$. Denoting the distribution of T_1 by $g(T_1)$ we have $\int_{-\infty}^{0} e^{\alpha T_1} g(T_1)\,dT_1 = M(\alpha\cdot 0)$. Change T_1 to $-T_1$. Since the probability that $-T_1$ lies in the range $-T_1 \mp \frac{1}{2}dT_1$ is $g(T_1)\,dT_1$,

$$\int_0^{\infty} e^{-\alpha T_1} g(T_1)\,dT_1 = M(\alpha\cdot 0) = \left(1+\frac{\alpha}{n\psi_1}\right)^{-n/2}\exp\left\{\frac{-\alpha\psi_2^2}{4\psi_1(\psi_1+\alpha/n)}\right\}$$

By Mellin's theorem, we have formally,

$$g(T_1) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\left(1+\frac{z}{n\psi_1}\right)^{-\frac{n}{2}}\exp\left\{zT_1 - \frac{z\psi_2^2}{4\psi_1(\psi_1+z/n)}\right\}dz \qquad (16)$$

We deal with this by the method of Section 6.3.2. First we note that, if we write $u(z)$ for the integrand, and put $k^2 = \psi_2^2/\psi_1^2$ and $z = r(\cos\theta + i\sin\theta)$, then

$$|z\,u(z)| = |z|\cdot|\exp zT_1|\cdot\left|\exp\frac{-z k^2}{4(1+z/n\psi_1)}\right|\cdot\left|1+z/n\psi_1\right|^{-n/2}$$

$$= r e^{T_1 r\cos\theta}\left\{\exp-\frac{k^2\psi_1 r(r+n\psi_1\cos\theta)}{4(r^2+2nr\psi_1\cos\theta+n^2\psi_1^2)}\right\}\left\{\begin{matrix}r^2+2nr\psi_1\cos\theta+n^2\psi_1^2\\ \times(n\psi_1)^{n/2}\end{matrix}\right\}^{-\frac{n}{2}}$$

Consider the behaviour of this expression when $r\to\infty$ and when $\frac{\pi}{2}\leq\theta\leq\frac{3\pi}{2}$. The index of the third factor is essentially negative, as $r\to\infty$ so this factor is

$$\leq\exp\left\{\frac{-k^2\psi_1 r(r-n\psi_1)}{4(r^2+n^2\psi_1^2)}\right\}\quad\text{when } \frac{\pi}{2}\leq\theta\leq\frac{3\pi}{2}$$

$$\leq\exp\left\{-\frac{k^2\psi_1}{4}\right\}\left\{\exp\frac{k^2 n\psi_1^2}{4r}\cdot\frac{1+n\psi_1/r}{1+n^2\psi_1^2/r^2}\right\}$$

$$= O\left(e^{1/r}\right)\quad\text{as } r\to\infty.$$

Hence

$$|z\,u(z)| < r\left\{\exp-\frac{k^2 n\psi_1^2}{4r}\cdot\frac{1+n\psi_1/r}{1+n^2\psi_1^2/r^2}\right\}(r-n\psi_1)^{-n}(n\psi_1)^{\frac{n}{2}}e^{-\frac{k^2\psi_1}{4}}$$

for all values of θ in the range $(\pi/2,\ 3\pi/2)$.

$$\therefore\ |z\,u(z)| = O\left(r^{-n+1}e^{1/r}\right)\longrightarrow 0$$

uniformly, as $r\to\infty$, with respect to θ when $\frac{\pi}{2}\leq\theta\leq\frac{3\pi}{2}$.

This result enables us to replace the integral in (16) by $\int_\Gamma u(z)\,dz$ where Γ is a contour formed by the line $c\pm i\infty$ and by the infinite semi-circle in the third and fourth quadrants. ($c >$ real part of any singularity of $u(z)$) The integrand has a singularity at $z=-n\psi_1$. To investigate its nature, we put $z=-n\psi_1+z'$ whence

$$u(z) = (n\psi_1)^{\frac{n}{2}} e^{-n\psi_1 T_1 - \frac{n\psi_2^2}{4\psi_1}} z'^{-\frac{n}{2}} e^{z'T + \frac{n^2\psi_2^2}{4z'}}$$

$$= (n\psi_1)^{\frac{n}{2}} e^{-n\psi_1 T_1 - \frac{n\psi_2^2}{4\psi_1}}$$

$$\times \left[1 + z' T_1 + \cdots \quad + z'^k T_1^k / k! + \cdots \right]$$

$$\times \left[\frac{1}{z'^{\frac{n}{2}}} + \frac{n^2\psi_2^2}{4 z'^{\frac{n}{2}+1}} + \cdots + \frac{n^{2k}\psi_2^{2k}}{4^k z'^{\frac{n}{2}+k}} \frac{1}{k!} + \cdots \right] \tag{17}$$

The two infinite series are absolutely convergent, so we may multiply them and rearrange the terms. We find, <u>if n is even</u>, that $u(z)$ is of the form

$$\cdots \frac{a_{-l}}{z'^{l}} + \frac{a_{-l+1}}{z'^{l-1}} + \cdots + \frac{a_{-1}}{z'} + a_0 + a_1 z' + \cdots + a_k z'^k \cdots \tag{18}$$

i.e., we find the Laurent expansion of $u(z)$, which is valid everywhere except at $z' = 0$, which point is an isolated essential singularity.

We must now prove that the Theorem of Residues applies to functions possessing isolated essential singularities. (In the text books, attention is usually confined to functions with non-essential singularities). Consider $\int_{\Gamma'} u(z)\,dz$ where Γ' denotes (i) the contour Γ, defined already, plus (ii) a circle Γ_1 of arbitrarily small radius δ surrounding $z' = 0$. Within the region bounded by Γ and Γ_1, $u(z)$ has no singularities. Therefore, by Cauchy's theorem, $\int_\Gamma u(z)\,dz = \int_{\Gamma_1} u(z)\,dz$; the integrals being taken in the same sense.

On $|z'| = \delta$, $u(z)$ is everywhere finite; so on this circle $u(z)$ is bounded for all values of $\arg z'$; consequently (18) is uniformly convergent with respect to $\arg z'$ on this contour, and may be integrated term by term. We find, as in the usual theory of contour integration (see e.g., Whittaker and Watson, 4th edition, p.111)

$$\int_{\Gamma_1} z'^{n} dz' = 0 \qquad \text{except when } n = -1$$

$$\int_{\Gamma_1} z'^{-1} dz' = 2\pi i,$$

Hence $\int_{\Gamma_1} u(z)\,dz = 2\pi i\, a_{-1}$, from (18)

and therefore $\int_{\Gamma} u(z)\,dz = 2\pi i\, a_{-1}$. By (16), also,

$$g(T_1) = a_{-1}, \quad \text{when } n \text{ is even.} \tag{19}$$

To evaluate a_{-1}, we return to (17);

$$\begin{aligned} a_{-1} &= \text{coefficient of } z'^{-1} \text{ in (17)} \\ &= (n\psi_1)^{\frac{n}{2}} e^{-n\psi_1 T_1 - \frac{n\psi_2^2}{4\psi_1}} \\ &\quad \times \left[\frac{T_1^{k-1}}{(k-1)!} + \frac{n^2\psi_2^2}{4} \frac{T_1^{k}}{k!} + \cdots + \frac{1}{r!}\left(\frac{n^2\psi_2^2}{4}\right)^r \frac{T_1^{k+r-1}}{(k+r-1)!} \cdots \right] \end{aligned}$$

(where $k = n/2$)

$$= (n\psi_1)^{\frac{n}{2}} e^{-n\psi_1 T_1 - \frac{n\psi_2^2}{4\psi_1}} \left(\frac{n\psi_2}{2}\right)^{-k+1} T_1^{\frac{k-1}{2}} I_{k-1}\left(n\psi_2 T_1^{1/2}\right) \tag{20}$$

where

$$I_{k-1}\left(n\psi_2 T_1^{1/2}\right) = \sum_{r=0}^{\infty} \left(\frac{n\psi_2 T_1^{1/2}}{2}\right)^{k-1+2r} \cdot \frac{1}{r!\,(k-1+r)!}$$

is Bessel's Function with imaginary argument (Whittaker and Watson 4th edition, p.372). Thus $I_{k-1}(z) = i^{-k+1} J_{k-1}(iz)$ in terms of the usual Bessel coefficient).

Reverting to the usual parameters m and σ^{2} instead of ψ_1 and ψ_2, we have, from (19) and (20)

$$g(T_1) = \frac{n}{2\sigma^2} m^{-\frac{n}{2}+1} e^{-2n m^2 \sigma^{-2}} T_1^{\frac{n-2}{4}} e^{-\frac{nT_1}{2\sigma^2}} I_{\frac{n-2}{2}}\left(\frac{n m T_1^{1/2}}{\sigma^2}\right)$$

where

$$T_1 = -\frac{1}{n} \sum x^2$$

So far, we know merely that this holds when n is even. Let us assume now that it is true for any integral value of n, and calculate the corresponding M.G.F.

In the general case, $I_{\frac{n-2}{2}}(z)$ is defined as

$$\sum_{r=0}^{\infty} z^{\frac{n-2}{2}+2r} \frac{1}{r!\,\Gamma\left(\frac{n}{2}+r\right)} \quad .$$

The r^{th} term in this sum is less than the r^{th} term in the expansion of $exp\left(z^{\frac{n-2}{2r}+2}\right)$. Since the latter is absolutely and uniformly convergent for all values of z, so too is $I_{\frac{n-2}{2}}(z)$. In evaluating the M.G.F., it is therefore permissible to integrate term by term. We have

$$\int_{-\infty}^{0} e^{\alpha T_1} g(T_1)\, dT_1$$
$$= (n\psi_1)^{n/2} e^{-\frac{n\psi_2^2}{4\psi_1}} \left(\frac{n^2\psi_2^2}{4}\right)^{-\frac{n-2}{4}}$$
$$\times \int_0^{\infty} \left\{ \sum_{r=0}^{\infty} \left(\frac{n\psi_2 T_1^{1/2}}{2}\right)^{\frac{n-2}{2}+2r} \frac{1}{r!\,\Gamma\left(\frac{n}{2}+r\right)} \right\} T_1^{\frac{n-2}{4}} e^{-(\alpha+n\psi_1)T_1}\, dT_1$$

The integration yields a sum of Gamma-functions, and we find, on reduction,

$$\int_{-\infty}^{0} e^{\alpha T_1} g(T_1)\, dT_1 = (n\psi_1)^{\frac{n}{2}} e^{-\frac{n\psi_2^2}{4\psi_1}} (\alpha+n\psi_1)^{-\frac{n}{2}} \sum_{r=0}^{\infty} \frac{1}{r!} \left\{\frac{n\psi_2^2}{4(\psi_1+\alpha/n)}\right\}^r$$
$$= \left(1+\frac{\alpha}{n\psi_1}\right)^{-\frac{n}{2}} exp\left\{\frac{-\alpha\,\psi_2^2}{4\psi_1(\psi_1+\alpha/n)}\right\}$$
$$= M(\alpha\cdot 0) .$$

Consequently (20) is the unique probability distribution of $-\frac{1}{n}\sum x^2$ for all values of n.

We may note in passing that (20) is of Koopman's form,

$$\exp\left\{ \alpha f(T_1) + F(\alpha) + c(T_1) \right\}$$

with

$$\alpha = n\psi_1$$

$$F(\alpha) = \frac{n}{2}\log\alpha - \frac{n^2\psi_2^{\,2}}{4\alpha} - \left(\frac{n}{2}-1\right)\log\frac{n\psi_2}{2}$$

$$c(T_1) = \frac{n-2}{4}\log T_1 + \log I_{\frac{n-2}{2}}\left(n\psi_2 T_1^{\,1/2}\right).$$

This is in accordance with the general theorem of Section 6.1.6.

CHAPTER SEVEN

NON-MINIMAL UNBIASSED STATISTICS

7.0 We may, in certain problems, be interested in the estimate of the variance of $T = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$ in the distribution

$$\varphi = \exp\{\psi f(x) + F(\psi) + c(x)\}.$$

This estimate of the variance is of course

$$\frac{1}{n}\sum_{i=1}^{n}\left\{f(x_i) - \left(-\frac{dF}{d\psi}\right)\right\}^2 = \frac{1}{n}\sum_{i=1}^{n} f^2(x_i) + \left(\frac{dF}{d\psi}\right)^2 + 2\frac{dF}{d\psi}.T$$

We may, furthermore, be concerned not merely with the mean (which we have already studied) but also with the variance, or higher moments, of this estimate. We are thus obliged to consider functions such as $\frac{1}{n}\sum_{i=1}^{n} f^2(x_i)$, $\frac{1}{n}\sum_{i=1}^{n} f^3(x_i)$, and, in general $\frac{1}{n}\sum_{i=1}^{n} f^r(x_i)$. Let us denote these functions by $T_{(2)}, T_{(3)}, \cdots T_{(r)}$ respectively.

7.1 The Mean and Variance of $T_{(r)}$: The M.G.F. of $T_{(r)}$ is

$$M(\alpha) = \int\cdots\int e^{\alpha T_{(r)}}\,\Phi\,dx'$$

$$= \left[\int \exp\left\{\frac{\alpha}{n} f^r(x) + \psi f(x) + F(\psi) + c(x)\right\}dx\right]^n$$

$$= \left[\int e^{\alpha f^r(x)/n}\,\varphi\,dx\right]^n$$

$$= \left[\int\left\{1 + \frac{\alpha f^r(x)}{n} + \frac{\alpha^2 f^{2r}(x)}{2!\,n^2} + \cdots\right\}\varphi\,dx\right]^n$$

$$= \left[1 + \frac{\alpha}{n}\nu'_r + \frac{\alpha^2}{2n^2}\nu'_{2r} + \cdots\right]^n$$

where $\nu'_p = \int f^p(x)\,\varphi\,dx$ (1)

$\therefore\ M(\alpha) = 1 + \alpha\nu'_r + \frac{\alpha^2}{2n}\left(\nu'_{2r} - \overline{n-1}\,\nu'^2_r\right) + O(\alpha^3)$

Therefore

(i) the mean of $T_{(r)}$ is ν'_r. (2)

(ii) the variance of $T_{(r)}$ about its mean is

$$\frac{1}{n}\left(\nu'_{2r} - \overline{n-1}\,\nu'^2_r\right) - \nu'^2_r = \frac{\nu'_{2r} - \nu'^2_r}{n} \tag{3}$$

The expressions (2) and (3) can be evaluated explicitly in terms of the elements of the original distribution by means of (1). In fact, ν'_p is the coefficient of $\beta^p/p!$ in the expansion of

$$\int e^{\beta f(x)}\varphi\,dx = \int e^{(\beta+\psi)f(x) + F(\psi) + c(x)}\,dx$$

$$= e^{F(\psi) - F(\psi+\beta)}\int e^{(\beta+\psi)f(x) + F(\psi+\beta) + c(x)}\,dx$$

$$= \exp\{F(\psi) - F(\psi+\beta)\} \tag{4}$$

(by the total probability condition). This last formula is, of course, merely a special case of equation (1) of Chapter Six. It follows, from the last Chapter or directly, that

$$\nu'_p = e^{F(\psi)}\frac{d^p e^{-F(\psi)}}{d\psi^p}$$

The terms ν'_p may be named the "natural moments" of φ about the origin.

Clearly, the mean of $T_{(r)}$ is independent of n, the number in the sample. Therefore $T_{(r)}$ is an <u>unbiassed</u> <u>estimate</u> of the coefficient $e^{F}\left(d^{r}e^{-F}/d\psi^{r}\right)$. We thus have found a whole family of unbiassed statistics for the Koopman distribution, since we can choose $r = 1, 2, 3, \cdots$ Also, functions of these statistics, say $\sum_r a_r T_{(r)}$ (a's arbitrary constants) are unbiassed.

<u>7.1.1</u> It is instructive to consider how $T_{(r)}$ can be expressed in terms of φ and of operations on φ. For example, when $r = 2$, we have

$$\log \varphi(x_i|\psi) = \psi f(x) + F(\psi) + c(x)$$

$$\therefore \frac{\partial \log \varphi(x_i|\psi)}{\partial \psi} = f(x) + \frac{dF}{d\psi}$$

$$\therefore \quad T_{(2)} = \frac{1}{n}\sum_i f^2(x_i) = \frac{1}{n}\sum_i \left(\frac{\partial \log \varphi(x_i|\psi)}{\partial \psi} - \frac{dF}{d\psi}\right)^2$$

which gives, on simplification,

$$T_{(2)} = -\frac{d^2F}{d\psi^2} + \left(\frac{dF}{d\psi}\right)^2 - \frac{2}{n}\frac{dF}{d\psi}\frac{\partial \log \Phi}{\partial \psi} + \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\varphi(x_i|\psi)}\frac{\partial^2 \varphi(x_i|\psi)}{\partial \psi^2} \qquad (5)$$

We recognise that (5) is a particular case of the general form of unbiassed statistic studied in Chapter Three. In that chapter, we learned that the necessary and sufficient condition for the stationary variance of the statistic was the vanishing of the terms in $\partial^2\varphi/\partial\psi^2$, $\partial^3\varphi/\partial\psi^3$, $\cdots$. In other words, $T_{(r)}$ though unbiassed, does not possess minimum variance except when $r = 1$. Similarly, $\sum_r a_r T_{(r)}$ is a non-minimal unbiassed statistic unless $a_r = 0,\ r \neq 0$.

<u>7.1.2</u> <u>Covariance of $T_{(r)}$ and T</u> :- The bivariate M.G.F. of $T_{(r)}$ and T (or T_1) $= \frac{1}{n}\sum_i f(x_i)$ is

$$M(\alpha, \beta) = \int \cdots \int e^{\alpha T + \beta T_{(r)}} \Phi \, dx'$$

$$= \int \cdots \int \exp\left\{ \frac{\alpha}{n} \sum_i f(x_i) + \frac{\beta}{n} \sum_i f^r(x_i) + \psi \sum_i f(x_i) + nF(\psi) + \sum_i c(x_i) \right\} dx'$$

$$= \left[\int \exp\left\{ \frac{\alpha f(x)}{n} + \frac{\beta f^r(x)}{n} \right\} \varphi \, dx \right]^n$$

$$= \left[1 + \frac{\alpha}{n} \nu_1' + \frac{\beta}{n} \nu_r' + \frac{\alpha\beta}{n^2} \nu_{r+1}' + \frac{\alpha^2}{2n^2} \nu_2' + \frac{\beta^2}{2n^2} \nu_{2r}' + \cdots \right]^n$$

in terms of the natural moments.

The covariance of $T_{(r)}$ and T is

= (coefficient of $\alpha\beta$) - (coefficient of α) $\times$ (coefficient of β)

$$= \left(\nu_{r+1}' - \nu_1' \nu_r' \right) / n \qquad (6)$$

Since (6) is not in general zero, $T_{(r)}$ and T are correlated. We can of course attempt to find, by means of this equation, the Koopman distribution for which T is uncorrelated with $T_{(r)}$ for some particular value of r. For instance, T and $T_{(2)}$ are uncorrelated if

$$\nu_3' - \nu_1' \nu_2' = 0$$

The natural moments are given by

$$\nu_1' = -dF/d\psi \; ; \qquad \nu_2' = -d^2F/d\psi^2 + (dF/d\psi)^2$$

$$\nu_3' = -d^3F/d\psi^3 + 3\left(d^2F/d\psi^2\right)\left(dF/d\psi\right) - \left(dF/d\psi\right)^3$$

so that our equation for zero covariance becomes

$$d^3F/d\psi^3 = 2\left(d^2F/d\psi^2\right)\left(dF/d\psi\right)$$

the integral of which is

$$F = \log\left\{ d \sec a(\psi + b) \right\}$$

where a, b, d are constants. Consequently, for the distribution

$$\varphi = \exp\left\{\psi f(x) + \log\left(d \sec a\sqrt{\psi+b}\right) + c(x)\right\}$$

the statistics $T = \frac{1}{n}\sum_i f(x_i)$ and $T_{(2)} = \frac{1}{n}\sum_i f^2(x_i)$

are uncorrelated in samples of n.

7.2 **Example**:- Consider the Koopman distribution

$$\varphi = \psi e^{-x\psi} \qquad (x \geqslant 0)$$

for which, in the usual notation, $f(x) = -x$; $F(\psi) = \log\psi$; $c(x) = 0$.

The M.G.F. of the natural moments about the origin is

$$M(\beta) = \exp\left\{F(\psi) - F(\psi+\beta)\right\} = \left(1 + \frac{\beta}{\psi}\right)^{-1}.$$

The r^{th} natural moment about the origin is therefore

$$\nu_r' = (-1)^r\, r!\, \psi^{-r}$$

By formulae (1) and (2), the mean of $T_{(r)} = \frac{1}{n}\sum(-x)^r$ is $\nu_r' = (-1)^r\, r!\, \psi^{-r}$, and the variance of $T_{(r)}$ is

$$V_{(r)} = \frac{1}{n}\left(\nu_{2r}' - \nu_r'^2\right) = \left\{(2r)! - (r!)^2\right\} / n\psi^{2r}.$$

We note that the variance of $T_{(r)}$ exceeds that of $T_{(r-1)}$ provided

$$V_{(r)} > V_{(r-1)}$$

i.e., provided

$$\psi^2 < \frac{(2r)! - (r!)^2}{(2r-2)! - (\overline{r-1}!)^2}$$

Hence $V_{(r)}$ <u>eventually</u> increases steadily with r, whatever the value of ψ. However, if ψ is known to have some particular value, it may be possible to determine a value r_0 of r for which the foregoing inequality is untrue. In this case, we should obtain a smaller absolute variance by choosing the r_0^{th} natural moment as an estimator, rather than the lower natural moments ν'_{r_0-1}, ν'_{r_0-2}, ... The result is of no importance in practice, of course, since r_0 depends on the population value of ψ, which is unknown.

Although the variance $V_{(r)}$ of $T_{(r)}$ may not always increase monotonically with r for small values of the latter, the coefficient of variation, say $v_{(r)}$ does increase steadily with r. By definition,

$$v^2_{(r)} = V_{(r)} \div \{\text{mean of } T_{(r)}\}^2$$
$$= \{(2r)! - (r!)^2\} / n(r!)^2$$

Therefore $v_{(r)} > v_{(r-1)}$ provided $2r\,(2r-1) > r^2$

i.e., provided $r > 2/3$, which is true. $[\,r = 1, 2, 3, \cdots\,]$

7.2.1 For the special case of $r=2$, the last Section shows that $T_{(2)} = \frac{1}{n} \sum x^2$ is an unbiassed estimate of $2/\psi^2$, with variance $20/n\psi^4$. Let us now apply the method of Maximum Likelihood to the estimation of this same coefficient.

Writing $\theta = 2/\psi^2$ we have

$$\varphi = \psi e^{-x\psi} = \sqrt{2/\theta} \cdot e^{-\sqrt{2}\cdot x/\sqrt{\theta}}$$

whence the likelihood function, in a sample of n, is

$$L = -\sqrt{\frac{2}{\theta}} \sum_i x_i + \frac{n}{2} \log 2 - \frac{n}{2} \log \theta .$$

$$\therefore \frac{\partial L}{\partial \hat{\theta}} = 0 \quad \text{gives } \frac{\sqrt{2}}{2} \hat{\theta}^{-\frac{3}{2}} \sum_i x_i - \frac{n}{2} \hat{\theta}^{-1} = 0$$

or $$\hat{\theta} = 2\left\{\frac{1}{n}\sum_i x_i\right\}^2,$$

which differs from our previous statistic $T_{(2)}$.

The variance, in large samples, of $\hat{\theta}$ is obtained from the formula

$$-1/\sigma_{\hat{\theta}}^2 = n\bar{b}$$

where

$$\begin{aligned}\bar{b} &= \int_0^\infty \varphi \frac{\partial^2 \log \varphi}{\partial \theta^2}\, dx \\ &= \int_0^\infty \left\{-\frac{3\sqrt{2}}{4\hat{\theta}^{-5/2}}\cdot x + \frac{1}{2\hat{\theta}^2}\right\}\frac{2}{\hat{\theta}^{1/2}} e^{-\sqrt{2}x/\hat{\theta}^{1/2}}\, dx \\ &= -\frac{1}{4}\hat{\theta}^{-2}.\end{aligned}$$

Therefore $\sigma_{\hat{\theta}}^2 = 4\hat{\theta}^2/n$, or, in terms of ψ,

$$\sigma_{\hat{\theta}}^2 = 16/n\psi^4.$$

The variance of the maximum likelihood estimate of $2/\psi^2$ is thus less, in large samples, than that of the alternative estimate $T_{(2)}$ - a result which was to be anticipated, since we know that the latter is non-minimal.

Does the maximum likelihood statistic $2\left(\sum x/n\right)^2$ possess a smaller variance than $T_{(2)}$ in small samples, we may ask? The M.G.F. of the former is

$$\begin{aligned}M'(\alpha) &= \int_0^\infty \cdots \int_0^\infty e^{2\alpha(\Sigma x/n)^2}\, \Phi\, dx' \\ &= 1 + 2\alpha\mu_2' + 2\alpha^2\mu_4' + O(\alpha^3)\end{aligned}$$

where μ_2', μ_4' are the second and fourth moments about the origin of $T = -\sum \frac{x}{n}$. (Formulae (1) and (3) of Chapter Six). For our distribution, in which $F(\psi) = \log \psi$, the M.G.F. of T is simply

$$\left(1 + \frac{\alpha}{n\psi}\right)^{-n}.$$

Hence $\mu_2' = \frac{1}{\psi^2}\left(1 + \frac{1}{n}\right)$; $\mu_4' = \frac{(n+1)(n+2)(n+3)}{n^3} \cdot \frac{1}{\psi^4}$

and we have the accurate results

(i) the mean of $2(\sum x/n)^2$ is $2\mu_2' = \frac{2}{\psi^2}\left(1 + \frac{1}{n}\right)$

(ii) the variance of this statistic is

$$\frac{4(\mu_4' - \mu_2'^2)}{n} = \frac{1}{n\psi^4}\left[16 + \frac{40}{n} + \frac{24}{n^2}\right]$$

which agrees with our previous expression $\sigma^2_{\hat{\theta}}$ only when n is so large that n^{-2} is negligible. Comparing this accurate value with the variance of $T_{(2)}$, we see that the latter is actually the smaller provided

$$\frac{20}{n\psi^4} < \frac{1}{n\psi^4}\left[16 + \frac{40}{n} + \frac{24}{n^2}\right]$$

i.e., provided $n \leq 10$.

For samples comprising 10 or fewer observations, therefore, the non-minimal unbiassed statistic $T_{(2)}$ has a smaller variance than the maximum likelihood estimate of the same parameter.

From the mean value quoted above, we see that $2(\sum x/n)^2$ is a biassed statistic in finite samples, in agreement with the theorem of Chapter Five (Section 5.5.1), which in this instance tells us that $\sum x/n$ is the only unbiassed statistic yielded by maximum likelihood.

7.2.2 Covariance of $T_{(2)}$ and $\hat{\theta}$:- The covariance of the alternative estimates $T_{(2)}$ and $\hat{\theta}$ of $2/\psi^2$ is

$$\rho = \int_0^\infty \cdots \int_0^\infty T_{(2)}\, \hat{\theta}\, \Phi\, dx' - \left(\int_0^\infty \cdots \int_0^\infty T_{(2)}\, \Phi\, dx'\right)\left(\int_0^\infty \cdots \int_0^\infty \hat{\theta}\, \Phi\, dx'\right).$$

Since, as we know

$$\int \cdots \int T_{(2)}\, \Phi\, dx' = 2/\psi^2$$

$$\int \cdots \int \hat{\theta}\, \Phi\, dx' = \frac{2}{\psi^2}\left(1 + \frac{1}{n}\right),$$

we have $\rho \quad = \int_0^\infty \cdots \int_0^\infty T_{(2)}\, \hat{\theta}\, \Phi\, dx' - \dfrac{4\left(1+\frac{1}{n}\right)}{\psi^4}$.

To evaluate this n-fold integral, we proceed as follows:-

$$I = \int_0^\infty \cdots \int_0^\infty T_{(2)}\, \hat{\theta}\, \Phi\, dx' = \int_0^\infty \cdots \int_0^\infty \left(\frac{1}{n}\sum_i x_i^2\right)\left(\frac{\sum x_i}{n}\right)^2 2\psi^n e^{-\psi \sum x_i}\, dx'$$

$$= \int_0^\infty \cdots \int_0^\infty \sum_{i=1}^{n} \frac{x_i^2}{n}\left\{\frac{x_i^2 + \sum' x_j^2 + 2x_i \sum' x_j}{n^2}\right\} 2\psi^n e^{-\psi x_i - \psi \sum' x_j}\, dx'$$

(where $\sum' x_j = \sum_{j=1}^{n} x_j - x_i$)

$$\therefore I = \int_0^\infty \cdots \int_0^\infty \frac{1}{n^2}\left[x^4 + 2x^3 \sum_{j=1}^{n-1} x_j + x^2 \sum_{j=1}^{n-1} x_j^2\right] 2\psi^n e^{-\psi x - \psi \sum_{j=1}^{n-1} x_j}\, dx\, dx_1 \cdots dx_{n-1}$$

$$= \frac{2\psi}{n^2}\int_0^\infty x^4 e^{-\psi x} dx \int \cdots \int \psi^{n-1} e^{-\psi \sum_1^{n-1} x_j}\, dx_1 \cdots dx_{n-1}$$

$$+ \frac{4\psi(n-1)}{n^2}\int_0^\infty x^3 e^{-\psi x} dx \int \cdots \int \psi^{n-1} \sum_{j=1}^{n-1} \frac{x_j}{n-1} \cdot e^{-\psi \sum_1^{n-1} x_j}\, dx_1 \cdots dx_{n-1}$$

$$+ \frac{2\psi(n-1)^2}{n^2}\int_0^\infty x^2 e^{-\psi x} dx \int \cdots \int \psi^{n-1}\left(\sum_{j=1}^{n-1} \frac{x_j}{n-1}\right)^2 e^{-\psi \sum_1^{n-1} x_j}\, dx_1 \cdots dx_{n-1}.$$

The single integrals in this expression are Gamma-functions. The $(n-1)$-fold integrals are the moments about the origin, of order 0, 1, 2 respectively, of the statistic $T = \frac{\Sigma x}{n-1}$ in samples of $n-1$, and are consequently given by the first three terms in the M.G.F.

$$\left(1 - \frac{\alpha}{(n-1)\psi}\right)^{-(n-1)}$$

i.e., they are 1; $1/\psi$; $\frac{n}{n-1}\frac{1}{\psi^2}$ respectively.

Hence, evaluating,

$$I = \frac{1}{n\psi^4}\left(4n + 20 + \frac{24}{n}\right)$$

and the covariance of $T_{(2)}$ and $\hat{\theta}$ is

$$\rho = \frac{16}{n\psi^4}\left(1 + \frac{3}{2n}\right)$$

7.3 The Variance of Biassed Maximum Likelihood Statistics:- In Section 7.2.1 we had occasion to verify the "maximum likelihood" expression for the variance of a particular biassed statistic yielded by this method of estimation. Let us digress for a moment to obtain the general verification of this result for the Koopman distribution.

We have seen that for

$$\varphi = \exp\{\psi f(x) + F(\psi) + c(x)\}$$

Maximum Likelihood gives $T = \frac{1}{n}\sum_i f(x_i)$ as an estimate of $-dF/d\psi$, and, say, $G(T)$ as an estimate - biassed - of $G(-dF/d\psi)$.

Suppose G can be expanded in a Maclaurin series

$$G(T) = \sum_j a_j T^j$$

The mean of G is $U = \int\cdots\int G\,\bar{\phi}\,dx' = \sum_j a_j \mu'_j$

where the μ'_j are the moments of T about the origin, i.e.,

$$\mu'_j = e^{nF(\psi)}\left(d^j e^{-nF(\psi)} / d\psi^j\right) \tag{7}$$

(We know of course that $\mu \to \sum_j a_j(-dF/d\psi)^j$ as $n \to \infty$)

The variance of G is

$$V = \int\cdots\int (g - \mu)^2 \Phi \, dx' = \int\cdots\int g^2 \Phi \, dx' - \mu^2$$

$$= \sum_j a_j^2 \left(\mu'_{2j} - \mu'^2_j\right) + \sum_{j \cdot k}{}' a_j \cdot a_k \left(\mu'_{j+k} - \mu'_j \mu'_k\right)$$

The notation $\sum'$ signifies that in the double summation, the terms $j = k$ are omitted.

To obtain V explicitly in terms of $F(\psi)$ and its derivatives, we use equation (7). For our purpose, it suffices to prove by induction that

$$\frac{d^j e^{-nF(\psi)}}{d\psi^j} = (-1)^j \left[n^j \left(\frac{dF}{d\psi}\right)^j - \frac{j(j-1)}{2!} n^{j-1} \frac{d^2F}{d\psi^2}\left(\frac{dF}{d\psi}\right)^{j-2} \right.$$

$$+ n^{j-2}\left\{ \frac{j(j-1)(j-2)}{3!} \frac{d^3F}{d\psi^3}\left(\frac{dF}{d\psi}\right)^{j-3} + \frac{j(j-1)(j-2)(j-3)}{2 \times 4}\left(\frac{d^2F}{d\psi^2}\right)^2\left(\frac{dF}{d\psi}\right)^{j-4} \right\}$$

$$\left. + O(n^{j-3}) \right] e^{-nF(\psi)} \tag{8}$$

Suppose (8) is true for a particular value of j.

Then $$\frac{d^{j+1} e^{-nF(\psi)}}{d\psi^{j+1}} = \frac{d}{d\psi} \text{(expression on right of (8))}$$

$$= (-1)^{j+1}\left[n^{j+1}\left(\frac{dF}{d\psi}\right)^{j+1} - \frac{(j+1)j}{2!}\, n^{j}\, \frac{d^2F}{d\psi^2}\left(\frac{dF}{d\psi}\right)^{j-1} \right.$$

$$+ n^{j-1}\left\{ \frac{(j+1)j(j-1)}{3!}\,\frac{d^3F}{d\psi^3}\left(\frac{dF}{d\psi}\right)^{j-2} + \frac{(j+1)j(j-1)(j-2)}{2\times 4}\left(\frac{d^2F}{d\psi^2}\right)^2\left(\frac{dF}{d\psi}\right)^{j-3}\right\}$$

$$\left. + O(n^{j-2}) \right] e^{-nF(\psi)}$$

on performing the differentiation and simplifying.

$=$ expression (8) with j replaced by $j+1$.

Since (8) can be verified at once for $j = 1, 2, 3, \cdots$ it is true in general. We utilise it to obtain the leading terms in V and find:

$$V = -\frac{d^2F}{n\,d\psi^2}\left\{ \sum_j j^2 a_j^{\,2}\left(\frac{dF}{d\psi}\right)^{2j-2} + 2\sum_{j\cdot k}{}' jk\,a_j a_k\left(\frac{-dF}{d\psi}\right)^{j+k-2}\right\}$$

$$+ \frac{1}{n^2}\left[\frac{d^3F}{d\psi^3}\left\{\sum_j j^2(j-1)a_j^{\,2}\left(\frac{dF}{d\psi}\right)^{2j-3} - 2\sum_{j\cdot k}{}' \frac{jk(j+k-2)}{2}\,a_j a_k\left(\frac{-dF}{d\psi}\right)^{j+k-3}\right\}\right]$$

$$+ \frac{1}{n^2}\left(\frac{d^2F}{d\psi^2}\right)^2\left\{\sum_j \frac{j^2(j-1)(3j-5)}{2}\,a_j^{\,2}\left(\frac{dF}{d\psi}\right)^{2j-4}\right.$$

$$\left. + 2\sum_{j\cdot k}{}' \frac{jk(j^2+jk+k^2-4j-4k+5)}{2}\,a_j a_k\left(\frac{-dF}{d\psi}\right)^{j+k-4}\right\}.$$

$$+ O(n^{-3})$$

Denoting the coefficient of n^{-2} by X we have

$$V = -\frac{1}{n}\frac{d^2F}{d\psi^2}\left\{\sum_j j a_j\left(\frac{-dF}{d\psi}\right)^{j-1}\right\}^2 + Xn^{-2} + O(n^{-3})$$

Consider, now, the approximation to V provided by the formula

$$-\frac{1}{\sigma_{\hat\theta}^2} = n\bar{b}$$

where

$$\theta = \sum_j a_j\left(-\frac{dF}{d\psi}\right)^j$$

and

$$\bar{b} = \int \varphi \frac{\partial^2 \log \varphi}{\partial \theta^2}\,dx$$

From the equation

$$\log \varphi = \psi f(x) + F(\psi) + c(x)$$

there results

$$\frac{\partial \log \varphi}{\partial \theta} = \frac{f(x) + dF/d\psi}{\sum_j j a_j\left(-\frac{dF}{d\psi}\right)^{j-1}\frac{d^2F}{d\psi^2}}$$

$$\frac{\partial^2 \log \varphi}{\partial \theta^2} = \left[\left\{ \sum_j -j\, a_j \left(-\frac{dF}{d\psi} \right)^{j-1} \left(\frac{d^2 F}{d\psi^2} \right)^2 \right\} - \left\{ f(x) + \frac{dF}{d\psi} \right\} \frac{d^2\theta}{d\psi^2} \right]$$

$$\times \left[\sum_j -j\, a_j \left(-\frac{dF}{d\psi} \right)^{j-1} \frac{d^2 F}{d\psi^2} \right]^{-3}$$

Remembering that

$$\int \varphi \, dx = 1 \quad ; \quad \int f(x) \varphi \, dx = -\frac{dF}{d\psi}$$

we have

$$\bar{b} = \left[\frac{d^2 F}{d\psi^2} \left\{ \sum_j j\, a_j \left(-\frac{dF}{d\psi} \right)^{j-1} \right\}^2 \right]^{-1}$$

so that

$$\sigma^2_{\hat{\theta}} = -\frac{1}{n} \frac{d^2 F}{d\psi^2} \left\{ \sum_j j\, a_j \left(-\frac{dF}{d\psi} \right)^{j-1} \right\}^2$$

Consequently

$$V - \sigma^2_{\hat{\theta}} = X n^{-2} + O(n^{-3})$$

This verifies the maximum likelihood rule for calculating the variance, since it gives a result differing from the true value V by an amount which is $O(n^{-2})$ in large samples.

7.4 Non-Minimal Unbiassed Statistics for the Two-Parameter Distribution:- Relative to the Koopman distribution

$$\varphi = \exp\left\{ \psi_1 f_1(x) + \psi_2 f_2(x) + F(\psi_1 \cdot \psi_2) + c(x) \right\}$$

let us calculate the mean and variance of the statistic

$$S = \frac{1}{n} \sum_i g\left[f_1(x_i), f_2(x_i) \right]$$

The M.G.F. of S is

$$M(\alpha) = \int \cdots \int e^{\alpha S} \Phi \, dx' = \left[\int e^{\frac{\alpha \mathcal{G}}{n}} \varphi \, dx \right]^n$$

$$= \left[1 + \frac{\alpha}{n} \int \mathcal{G} \varphi \, dx + \frac{\alpha^2}{2n^2} \int \mathcal{G}^2 \varphi \, dx + O(\alpha^3) \right]^n$$

Write

$$\int \mathcal{G} \varphi \, dx = \mu_1 \quad ; \quad \int \mathcal{G}^2 \varphi \, dx = \mu_2$$

Then $M(\alpha) = 1 + \alpha \mu_1 + \frac{\alpha^2}{2} \left(\frac{\mu_2}{n} + \frac{n-1}{n} \mu_1^2 \right) + O(\alpha^3)$

and

(i) the mean of S is μ_1 (9)

(ii) the variance of S is

$$\frac{1}{n} \left\{ \mu_2 + \overline{n-1}\, \mu_1^2 \right\} - \mu_1^2 = \frac{\mu_2 - \mu_1^2}{n} \tag{10}$$

Suppose that G is capable of expansion in a Maclaurin series, say

$$\mathcal{G} = \sum_{r.s} a_{r.s} f_1^r f_2^s \qquad (a\text{'s constants})$$

We can now evaluate μ_1, μ_2 in terms of the "bivariate natural moments" of the distribution. The $(r.s)^{th}$ bivariate natural moment is defined as

$$\nu'_{r.s} = \int f_1^r(x) f_2^s(x) \varphi \, dx \tag{11}$$

and is, consequently, the coefficient of $\alpha^r \beta^s / r! s!$ in

$$\int e^{\alpha f_1 + \beta f_2} \varphi \, dx = \int e^{(\alpha+\psi_1) f_1 + (\beta+\psi_2) f_2 + F(\psi_1 \cdot \psi_2) + c(x)} dx$$

$$= \exp \left\{ F(\psi_1 \cdot \psi_2) - F(\psi_1 + \alpha \cdot \psi_2 + \beta) \right\}$$

In fact $\mu_1 = \int \mathcal{G} \varphi \, dx = \sum_{r.s} a_{r.s} \int f_1^r f_2^s \varphi \, dx = \sum_{r.s} a_{rs} \nu'_{rs}$ (12)

and $\mu_2 = \int \mathcal{G}^2 \varphi \, dx = \sum_{r.s} a_{r.s}^2 \nu'_{2r.2s} + 2 \sum'_{r.s.j.k} a_{rs} \cdot a_{j.k} \nu'_{r+j} \cdot \nu'_{s+k}$

(13)

where $\sum'$ denotes that the terms for which $r=j$, $s=k$ simultaneously are omitted.

Substituting (12) and (13) into equations (9) and (10) respectively, we obtain the mean and variance of S'. It is clear that (9) is independent of the size of the sample n i.e. S is an unbiassed statistic. By expressing S in terms of φ and derivatives of φ, it is readily appreciated that it is of the general form of unbiassed statistic considered in Chapter Five. It is therefore non-minimal, unless all the constants $a_{r,s}$ are zero except a_{10} and a_{01} i.e., apart from this case, S is never one of a pair of statistics which have minimum generalised variance.

7.4.1 Covariance of S and One of the Unbiassed Statistics of Minimum Generalised Variance:- The bivariate M.G.F. of S and, say, $T_1 = \frac{1}{n}\sum_i f_1(x_i)$ is

$$M(\alpha\cdot\beta) = \int\cdots\int e^{\alpha S + \beta T_1}\,\Phi\,dx'$$
$$= \left\{\int e^{\frac{\alpha g}{n} + \frac{\beta f_1}{n}}\varphi\,dx\right\}^n$$
$$= \left[\int\left\{1+\frac{\alpha g}{n} + O(\alpha^2)\right\}\left\{1+\frac{\beta f_1}{n} + O(\beta^2)\right\}\varphi\,dx\right]^n$$
$$= \left[1+\frac{\alpha}{n}\mu_1 + \frac{\beta}{n}\nu'_{10} + \frac{\alpha\beta}{n^2}\int f_1 g\varphi\,dx + \cdots\right]^n$$
$$= 1+\alpha\mu_1+\beta\nu'_{10} + \frac{\alpha\beta}{n}\left(\int f_1 g\varphi\,dx + \overline{n-1}\,\mu_1\nu'_{10}\right)+\cdots$$

The covariance is thus

$$\rho = \frac{1}{n}\left(\int f_1 g\varphi\,dx - \mu_1\nu'_{10}\right)$$

Now, since $g = \sum_{r.s} a_{r.s} f_1^r f_2^s$,

$$\int f_1\, g\, \varphi\, dx = \sum_{r.s} a_{r.s} V'_{r+1.s}$$

Hence, remembering (12),

$$\rho = \frac{1}{n} \sum_{r.s} a_{r.s} \left(V'_{r+1.s} - V'_{10} \cdot V'_{r.s} \right) \tag{14}$$

7.4.2 Examples:-

(a) Let us choose $g[f_1 \cdot f_2] \equiv f_2$. i.e., we choose $a_{01} = 1$ and all the other a's are zero. Equation (14) then gives the covariance of $S = \frac{1}{n} \sum_i f_2(x_i)$ and T_1 as

$$\rho = \frac{1}{n} \left(V'_{11} - V'_{10} V'_{01} \right).$$

The V's are the successive terms in the expansion of

$$\exp\left\{F(\psi_1 \cdot \psi_2) - F(\psi_1 + \alpha \cdot \psi_2 + \beta)\right\}$$
$$= 1 - \alpha \frac{\partial F}{\partial \psi_1} - \beta \frac{\partial F}{\partial \psi_2} + \frac{\alpha^2}{2}\left(-\frac{\partial^2 F}{\partial \psi_1^2} + \overline{\frac{\partial F}{\partial \psi_1}}^2 \right)$$
$$+ \frac{\beta^2}{2}\left(-\frac{\partial^2 F}{\partial \psi_2^2} + \overline{\frac{\partial F}{\partial \psi_2}}^2 \right) + \alpha\beta\left(-\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} + \frac{\partial F}{\partial \psi_1} \cdot \frac{\partial F}{\partial \psi_2} \right) + \cdots$$

so that

$$V'_{10} = -\frac{\partial F}{\partial \psi_1}\ ; \quad V'_{01} = -\frac{\partial F}{\partial \psi_2}\ ; \quad V'_{11} = -\frac{\partial^2 F}{\partial \psi_1 \partial \psi_2} + \frac{\partial F}{\partial \psi_1} \cdot \frac{\partial F}{\partial \psi_2}$$

Consequently $\rho = -\frac{1}{n} \frac{\partial^2 F}{\partial \psi_1 \partial \psi_2}$, which is, of course, the correct result.

(b) Choose $g[f_1 \cdot f_2] \equiv f_1$. i.e., take $a_{10} = 1$, and all the other a's zero. The covariance of S and T_1 is in this instance merely the variance of T_1. Formula (14) gives

$$\rho = \frac{1}{n} \left(V'_{20} - V'^{\,2}_{10} \right)$$

and, from the preceding expansion, $V'_{20} = -\frac{\partial^2 F}{\partial \psi_1^2} + \left(\frac{\partial F}{\partial \psi_1}\right)^2$. Therefore $\rho = -\frac{1}{n}\frac{\partial^2 F}{\partial \psi_1^2}$, which, again, is correct.

7.5 Sufficient Statistics in Koopman's Sense and in Fisher's Sense:- On the numerous occasions of reference to sufficient statistics, we have always used the term in the sense of Koopman's definition, as quoted in Chapter Three. Koopman, we recall, claimed that he was making precise "the intuitive idea of such a statistic" - viz., a statistic which utilises all the information provided by the sample. However, other definitions of sufficiency have been given. R. A. Fisher himself propounded one in his original paper on Maximum Likelihood ("The Foundations of Theoretical Statistics," Philsophical Transactions A, Vol. 222, 1922) in the following terms.

"Let T be a statistic which estimates a coefficient θ, and let T^x be any other estimate of θ. T is sufficient if the joint sampling distribution of T and T^x is of the form

$$p(T, T^x|\theta) = p_1(T|\theta)\, p_2(T^x|T) \qquad (15)$$

where p_1 is the probability distribution of T in a population specified by the parameter θ; and where p_2 is the probability distribution of T^x in a population specified by the parameter T."

From an example in the paper referred to, it is possible that this definition was intended to apply only to statistics T, T^x whose simultaneous distribution is normal, or - virtually the same restriction - only to indefinitely large samples. We shall now show by an example that, in finite samples at least, Fisher's and Koopman's definitions are not equivalent.

7.5.1 Suppose that T, T^x are jointly distributed in the foregoing form, and consider the distribution p' of T and T' where T' is some function of T^x, say $T' = g(T^x)$ - thus T' is an estimate of $g(\theta)$.

Denoting the inverse of $T' = g(T^x)$ by $T^x = g^{-1}(T')$ we have

$$p' = p_1(T|\theta) \cdot p_2\left\{g^{-1}(T') \mid T\right\} \frac{dg^{-1}}{dT'}$$

or, writing $p_2\left\{g^{-1}(T') | T\right\} \frac{dg^{-1}}{dT'} = p_3(T'|T)$,

$$p' = p_1(T|\theta)\, p_3(T'|T).$$

which is of the form (15), even though T and T' are estimates of different coefficients, θ and $g(\theta)$.

Assuming that T and T' both vary between $\pm\infty$, the bivariate M.G.F. of these two statistics is

$$M(\alpha,\beta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\alpha T + i\beta T'}\, p'\, dT\, dT'$$

$$= \int_{-\infty}^{\infty} e^{i\alpha T} p_1(T|\theta)\, dT \int_{-\infty}^{\infty} e^{i\beta T'} p_3(T'|T)\, dT'.$$

Now $\int_{-\infty}^{\infty} e^{i\beta T'} p_3(T'|T)\, dT' = \text{M.G.F. of } T'$ in a population specified by the parameter T

$$= M_0(\beta|T) \text{ say}$$

whence $M(\alpha,\beta) = \int_{-\infty}^{\infty} e^{i\alpha T} p_1(T|\theta)\, M_0(\beta|T)\, dT$

p_1 and p_3, being probability distributions, are bounded functions. Therefore the coefficient of $e^{i\alpha T}$ in the foregoing integrand is bounded, and inversion of the integral by means of Fourier transform theory is permissible. This gives

$$p_1(T/\theta)\, M_0(\beta/T) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\alpha T} M(\alpha \cdot \beta)\, d\alpha .$$

This equation serves as a test, in the following manner. Let two statistics T, T', estimates of $\theta, g(\theta)$ respectively, be given. We calculate the sampling distribution of T, say $p_1(T/\theta)$. We further calculate the bivariate M.G.F. of T and T', say $M(\alpha \cdot \beta)$. Now if the Fourier transform of $M(\alpha \cdot \beta)$, divided by $p_1(T/\theta)$, is a function of θ, then T is certainly not sufficient in Fisher's sense.

7.5.2 Fundamentally, the proposed test involves the finding of a joint probability distribution from a given bivariate M.G.F., and the validity of the argument depends on whether the distribution so obtained is unique. When the range of both T and T' is $\pm\infty$, this condition is fulfilled; we merely invoke the uniqueness theorem of Fourier transforms of several variables (see e.g., Bochner, "Vorlesungen über Fouriersche Integrale"). For other ranges, the same result holds, so far as functions of class L^2 are concerned, by the theory of linear integral equations - Chapter Two.

7.5.3 Example:- For the distribution

$$\varphi = \exp\{\psi f(x) + F(\psi) + c(x)\}$$

the statistic $T = \frac{1}{n}\sum_i f(x_i)$ is sufficient, in Koopman's sense, for the estimation of $\theta = -dF/d\psi$. As the T' of Section 7.5.1, we choose the unbiassed, non-minimal, statistic $\frac{1}{n}\sum_i f^2(x_i)$.

The bivariate M.G.F. of T and T' is

$$M(\alpha\cdot\beta) = \int\cdots\int e^{i\alpha T + i\beta T'}\,\varphi\,dx'$$

$$= \left[\int \exp\left\{\frac{i\beta f^2}{n} + \left(\psi + \frac{i\alpha}{n}\right)f + F + c\right\}dx\right]^n$$

$$= \left[\sum_{r=0}^{\infty}\left(\frac{i\beta}{n}\right)^r\frac{1}{r!}\int f^{2r}\exp\left\{\left(\psi+\frac{i\alpha}{n}\right)f + F + c\right\}dx\right]^n$$

Now

$$\int \exp\left\{\left(\psi+\frac{i\alpha}{n}\right)f + F + c\right\}dx = \exp\left\{F(\psi) - F\left(\psi + \frac{i\alpha}{n}\right)\right\},$$

in consequence of the total probability condition $\int\varphi\,dx = 1$ for all values of ψ.

Differentiating r times with respect to α,

$$\int\left(\frac{if}{n}\right)^r \exp\left\{\left(\psi+\frac{i\alpha}{n}\right)f + F + c\right\}dx = \frac{d^r}{d\alpha^r}\exp\left\{F(\psi) - F\left(\psi+\frac{i\alpha}{n}\right)\right\}$$

$$= \frac{d^r M_0(\alpha)}{d\alpha^r}, \quad say.$$

Hence

$$M(\alpha\cdot\beta) = \left[\sum_{r=0}^{\infty}(-i)^r\frac{(n\beta)^r}{r!}\frac{d^{2r}M_0(\alpha)}{d\alpha^{2r}}\right]^n$$

To apply our test, we ascertain whether, for all values of β for which the foregoing expansion is valid,

$$\left(\int_{-\infty}^{\infty} e^{-i\alpha T} M(\alpha\cdot\beta)\,d\alpha\right)\Big/ p_1(T|\psi)$$

is independent of ψ. (We assume that the limits of T are $\pm\infty$). Our expression for $M(\alpha\beta)$ is a power series in β. Hence if the test fails for any term in this expansion, it follows that T is not sufficient in Fisher's sense.

The coefficient of β in $M(\alpha\cdot\beta)$ is

$$-in\frac{d^2M_0(\alpha)}{d\alpha^2} = -\frac{i}{n}\left[\left\{\frac{d^2F(\psi+\frac{i\alpha}{n})}{d\psi^2} - \left(\frac{dF(\psi+\frac{i\alpha}{n})}{d\psi}\right)^2\right\}e^{F(\psi)-F(\psi+\frac{i\alpha}{n})}\right]$$

Moreover, since the M.G.F. of T alone is $\exp\{F(\psi)-F(\psi+\frac{i\alpha}{n})\}$ we have

$$\int_{-\infty}^{\infty} e^{i\alpha T} p_1(T|\psi)\,dT = \exp\{F(\psi)-F(\psi+\tfrac{i\alpha}{n})\}$$

Therefore $p_1(T|\psi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\alpha T+F(\psi)-F(\psi+\frac{i\alpha}{n})}\,d\alpha$.

Our test can therefore be put in the form - is

$$\frac{\int_{-\infty}^{\infty}\left\{\frac{d^2F(\psi+\frac{i\alpha}{n})}{d\psi^2} - \left(\frac{dF(\psi+\frac{i\alpha}{n})}{d\psi}\right)^2\right\}\exp\{-i\alpha T+F(\psi)-F(\psi+\frac{i\alpha}{n})\}\,d\alpha}{\int_{-\infty}^{\infty}\exp\{-i\gamma T+F(\psi)-F(\psi+i\gamma/n)\}\,d\gamma} \tag{16}$$

independent of ψ? Alternatively, is

$\frac{d}{d\psi}$ (quotient above) $=0$ for all values of ψ and T?

Differentiating and simplifying, and writing $\psi+\frac{i\alpha}{n}=z$, $\psi+\frac{i\gamma}{n}=\omega$, we find that this is untrue unless

$$\int_{\psi-i\infty}^{\psi+i\infty}\int_{\psi-i\infty}^{\psi+i\infty}\exp\{-nzT-n\omega T-nF(z)-nF(\omega)\}\,dz\,d\omega \times\left[\frac{d^3F(z)}{dz^3}-2\frac{d^2F(z)}{dz^2}\frac{dF(z)}{dz} - n\left\{\frac{d^2F(z)}{dz^2}-\left(\frac{dF(z)}{dz}\right)^2\right\}\left\{\frac{dF(z)}{dz}-\frac{dF(\omega)}{d\omega}\right\}\right] = 0$$

for all values of ψ and T

Putting

$$\mu(z\cdot\omega) = \exp\{-nF(z) - nF(\omega)\}$$
$$\times\left[\frac{d^3F(z)}{dz^3} - 2\frac{d^2F(z)}{dz^2}\cdot\frac{dF}{dz} - n\left\{\frac{d^2F(z)}{dz^2} - \overline{\frac{dF}{dz}}^2\right\}\left\{\frac{dF(z)}{dz} - \frac{dF(\omega)}{d\omega}\right\}\right] \quad (17)$$

this condition becomes

$$\int_{\psi-i\infty}^{\psi+i\infty}\int_{\psi-i\infty}^{\psi+i\infty} \exp\{-nzT - n\omega T\}\,\mu(z\cdot\omega)\,dz\,d\omega = 0$$

for all values of ψ and T. The left hand side of this expression is the Laplace transform of the function μ of two variables, and the condition that this transform be always zero is (Bochner, loc. cit.)

$$\mu(z\cdot\omega) \equiv 0 .$$

Inspecting (17) we note that this identity implies

$$\frac{d^2F(z)}{dz^2} - \left(\frac{dF(z)}{dz}\right)^2 = 0$$

or (integrating)

$$F(z) = -\log A(z+a)$$

where A, a are constants. If F is not of this restricted form, the foregoing transform is not identically zero; therefore the quotient (16) is dependent on ψ. It follows that the statistic $T = \frac{1}{n}\sum_i f(x_i)$, which is sufficient in Koopman's sense, is not sufficient in Fisher's sense.